AD_____

Award Number: W81XWH-07-1-0352

TITLE: Site-Specific Differentiation of Fibroblasts in Normal and Scleroderma Skin

PRINCIPAL INVESTIGATOR: Howard Y. Chang, M.D., Ph.D.

CONTRACTING ORGANIZATION: Leland Stanford Junior University
Stanford, CA 94305-4125

REPORT DATE: June 2008

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-06-2008	Annual	1 JUN 2007 - 31 MAY 2008

4. TITLE AND SUBTITLE

Site-Specific Differentiation of Fibroblasts in Normal And Scleroderma Skin

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-07-1-0352

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Howard Y. Chang, M.D., Ph.D.

E-Mail: howchang@stanford.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Leland Stanford Junior University
Stanford, CA 94305-4125

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The results from year 1 of funding suggest that appropriate markers and technologies are in place for systematic investigation of the positional identity of fibroblasts, both in normal and diseased tissues. Because long noncoding RNAs are the predominant position specific output of the HOX loci, they may constitute excellent markers for testing the idea that scleroderma involves altered fibroblast positional identity.

15. SUBJECT TERMS
No subject terms provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	26	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

	Page
Introduction……..	**1**
Body……..	**1**
Key Research Accomplishments……..	**2**
Reportable Outcomes……	**2**
Conclusion……	**3**
References……	**3**
Appendices……	**4-24**

INTRODUCTION:

In systemic sclerosis (SSc), fibrosis of skin and internal organs occurs in an anatomically reproducible and progressive fashion. The extent and localization of skin involvement are important predictors of long-term outcome and mortality, but the basis of site specificity in SSc is not understood. Fibroblasts are the principal cells in dermis and stroma of epithelial organs that synthesize extracellular matrix proteins, and they are believed to be responsible for excessive fibrosis and tissue hardening in SSc. We recently discovered that fibroblasts are systematically differentiated in a site-specific manner, and they vary significantly in the expression of many genes related to extracellular matrix synthesis and turnover. We hypothesize that the site-specific differentiation of fibroblasts plays an important role in the anatomic specificity of SSc, addressed in three specific aims. At the end of the funding period, we will have characterized fibroblast gene expression patterns in both normal and scleroderma skin and identified the specific fibroblast populations and their gene products that might be associated with SSc progression. This new knowledge will provide basic and much-needed insights on fibroblasts and the pathogenesis of SSc.

Specific Aim 1: Profile fibroblast populations from skin of diverse anatomic sites, thereby creating a reference database of markers for site-specific fibroblast populations.

Specific Aim 2: Construct tissue microarrays of skin and scleroderma tissues to enable high throughput visualization of fibroblast gene and protein expression.

Specific Aim 3: Quantify site-specific populations of fibroblasts in SSc and compare them with site-matched normal and control skin tissues

BODY:

To test the hypothesis that altered fibroblast positional identity may relate to SSc pathogenesis, our first task is to identify reliable markers of fibroblast positional identity in normal skin tissues. While previous works have examined the variation in protein coding genes, it is now recognized that noncoding transcription, yielding regulatory RNAs that do not generate protein, is a prominent feature of the human genome (Amaral et al., 2008; Kapranov et al., 2007). Noncoding RNAs (ncRNA) may be as numerous and diverse in a given tissue as protein coding genes. The HOX loci encode genes important for skin positional identity and have noncoding transcription (Sessa et al., 2007), but the identity, expression pattern, and scope of ncRNA transcription were unclear. Thus, we characterize the transcriptional landscape of the four human HOX loci at five base pair resolution in 11 anatomic sites [(Rinn et al., 2007), manuscript attached in Appendix]. We identified 231 HOX ncRNAs that extend known transcribed regions by more than 30 kilobases. HOX ncRNAs are spatially expressed along developmental axes and possess unique sequence motifs, and their expression demarcates broad chromosomal domains of differential histone methylation and RNA polymerase accessibility. We identified a 2.2 kilobase ncRNA residing in the HOXC locus, termed HOTAIR, which represses transcription in trans across 40 kilobases of the HOXD locus. HOTAIR interacts with Polycomb Repressive Complex 2 (PRC2) and is required for PRC2 occupancy and histone H3 lysine-27 trimethylation of HOXD locus. Thus, transcription of ncRNA may demarcate chromosomal domains of gene silencing at a distance; these results have broad implications for gene regulation in development and disease states.

In the second aim of the proposal, we proposed to create tissue microarrays of skin in order to facilitate high throughput interrogation of RNA or protein expression. Gene and tiling microarrays are useful for elucidating the genetic and epigenetic elements that differentiate cell types across the body; however it is equally important to determine the *in vivo* anatomic localization of genes in the three-dimensional context of the skin. To address this challenge, we have constructed a “skin diversity” tissue microarray, where multiple skin sections are placed on a single slide to be used for in situ hybridization and immunohistochemistry. Our tissue microarrays is comprised of 42, two-millimeter formalin-fixed, paraffin-embedded cores of skin from diverse anatomic sites and 8 internal organs such as cervix, intestine, lung, liver and bone (Rinn et al., 2008). Immunohistochemistry or RNA in situ hybridization can be performed on all 50 tissues in parallel, allowing unbiased and high throughput comparison of protein or gene expression levels and localization. A potential limitation of this technology is that proteins and mRNAs present in low levels may be better visualized in frozen sections than formalin-fixed tissues, and conditions for antigen retrieval or signal amplification may need to be developed to visualize low abundance gene products.

To illustrate the use of such a skin diversity tissue microarray, we performed RNA in situ hybridizations for Keratin 14 (K14) and Keratin 9 (K9). As expected, we observed expression of K14 in the basal layer of epidermis in skin from all anatomic sites. K9 is a suprabasal keratin of plamo-plantar skin, and indeed only palmo-plantar skin on our tissue microarray showed strong K9 signal (Rinn et al., 2008). The skin diversity tissue microarray should be useful for the discovery and validation of novel site-specific genes or signaling pathways. Again, this approach can be extended to include tissues of skin diseases to monitor the in vivo expression of genes perturb in disease. The combined power of gene expression, tiling and tissue microarrays will greatly facilitate our understanding of the genes that are important in skin patterning and their roles in skin disease.

KEY RESEARCH ACCOMPLISHMENTS:

- Identification of new position-specific gene markers, including noncoding RNAs
- Construction of skin tissue microarrays
- Use of skin diversity TMA to interrogate site-specific gene expression

REPORTABLE OUTCOMES:

Two publications, listed below.

1. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

2. Rinn, J. L., Wang, J. K., Liu, H., Montgomery, K., van de Rijn, M., and Chang, H. Y. (2008). A systems biology approach to anatomic diversity of skin. J Invest Dermatol *128*, 776-782.

CONCLUSION:

These results suggest that appropriate markers and technologies are in place for systematic investigation of the positional identity of fibroblasts, both in normal and diseased tissues. Because long ncRNAs are the predominant position specific output of the HOX loci, they may constitute excellent markers for testing the idea that scleroderma involves altered fibroblast positional identity.

REFERENCES:

Amaral, P. P., Dinger, M. E., Mercer, T. R., and Mattick, J. S. (2008). The eukaryotic genome as an RNA machine. Science *319*, 1787-1789.

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermuller, J., Hofacker, I. L.*, et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-1488.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

Rinn, J. L., Wang, J. K., Liu, H., Montgomery, K., van de Rijn, M., and Chang, H. Y. (2008). A systems biology approach to anatomic diversity of skin. J Invest Dermatol *128*, 776-782.

Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G., and Orlando, V. (2007). Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. Rna *13*, 223-239.

APPENDICES:

p. 4-16

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

p. 17-24.

Rinn, J. L., Wang, J. K., Liu, H., Montgomery, K., van de Rijn, M., and Chang, H. Y. (2008). A systems biology approach to anatomic diversity of skin. J Invest Dermatol *128*, 776-782.

SUPPORTING DATA: None.

Cell

Functional Demarcation of Active and Silent Chromatin Domains in Human *HOX* Loci by Noncoding RNAs

John L. Rinn,[1] Michael Kertesz,[3,5] Jordon K. Wang,[1,5] Sharon L. Squazzo,[4] Xiao Xu,[1] Samantha A. Brugmann,[2] L. Henry Goodnough,[2] Jill A. Helms,[2] Peggy J. Farnham,[4] Eran Segal,[3,*] and Howard Y. Chang[1,*]
[1] Program in Epithelial Biology
[2] Department of Surgery
Stanford University School of Medicine, Stanford, CA 94305, USA
[3] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
[4] Department of Pharmacology and Genome Center, University of California, Davis, CA 95616, USA
[5] These authors contributed equally to this work.
*Correspondence: howchang@stanford.edu (H.Y.C.), eran.segal@weizmann.ac.il (E.S.)
DOI 10.1016/j.cell.2007.05.022

SUMMARY

Noncoding RNAs (ncRNA) participate in epigenetic regulation but are poorly understood. Here we characterize the transcriptional landscape of the four human *HOX* loci at five base pair resolution in 11 anatomic sites and identify 231 *HOX* ncRNAs that extend known transcribed regions by more than 30 kilobases. *HOX* ncRNAs are spatially expressed along developmental axes and possess unique sequence motifs, and their expression demarcates broad chromosomal domains of differential histone methylation and RNA polymerase accessibility. We identified a 2.2 kilobase ncRNA residing in the *HOXC* locus, termed HOTAIR, which represses transcription in trans across 40 kilobases of the *HOXD* locus. HOTAIR interacts with Polycomb Repressive Complex 2 (PRC2) and is required for PRC2 occupancy and histone H3 lysine-27 trimethylation of *HOXD* locus. Thus, transcription of ncRNA may demarcate chromosomal domains of gene silencing at a distance; these results have broad implications for gene regulation in development and disease states.

INTRODUCTION

A distinguishing feature of metazoan genomes is the abundance of noncoding RNA (ncRNAs), which function by means other than directing the production of proteins. In addition to small regulatory RNAs such as miRNAs, recent studies have predicted the existence of long ncRNAs—ranging from 300 nucleotides (nt) to over 10 kb—that are spliced, polyadenylated, and roughly as diverse in a given cell type as protein-coding mRNAs (Bertone et al., 2004; Carninci et al., 2005; Kapranov et al., 2005; Rinn et al., 2003). Long ncRNAs may have diverse roles in gene regulation, especially in epigenetic control of chromatin (Bernstein and Allis, 2005). Perhaps the most prominent example is silencing of the inactive X chromosome by the ncRNA XIST. To normalize the copy number of X chromosomes between male and female cells, transcription of XIST RNA from one of the two female X chromosome is involved in recruiting Polycomb group proteins (PcG) to trimethylate histone H3 on lysine 27 (H3K27me3), rendering the chromosome transcriptionally silent (Plath et al., 2003). It is believed that Polycomb Repressive Complex 2 (PRC2), comprised of H3K27 histone methyl transferase (HMTase) EZH2 and core components Suz12 and EED, initiates this histone modification and that, subsequently, Polycomb Repressive Complex 1 (PRC1) maintains this modification and promotes chromatin compaction (reviewed by Sparmann and van Lohuizen, 2006). Presently, the mechanism by which XIST ncRNA guides Polycomb activity is unclear. Several PcG proteins possess RNA-binding activity, and RNA is required for PcG binding to DNA, suggesting that specific ncRNAs may be critical interfaces between chromatin-remodeling complexes and the genome (Bernstein et al., 2006; Zhang et al., 2004).

In addition to dosage compensation, long ncRNAs may also play critical roles in pattern formation and differentiation. In mammals, 39 *HOX* transcription factors clustered on four chromosomal loci, termed *HOXA* through *HOXD*, are essential for specifying the positional identities of cells. The temporal and spatial pattern of *HOX* gene expression is often correlated to their genomic location within each loci, a property termed colinearity (Kmita and Duboule, 2003; Lemons and McGinnis, 2006). Maintenance of *HOX* expression patterns is under complex epigenetic regulation. Two opposing groups of histone-modifying

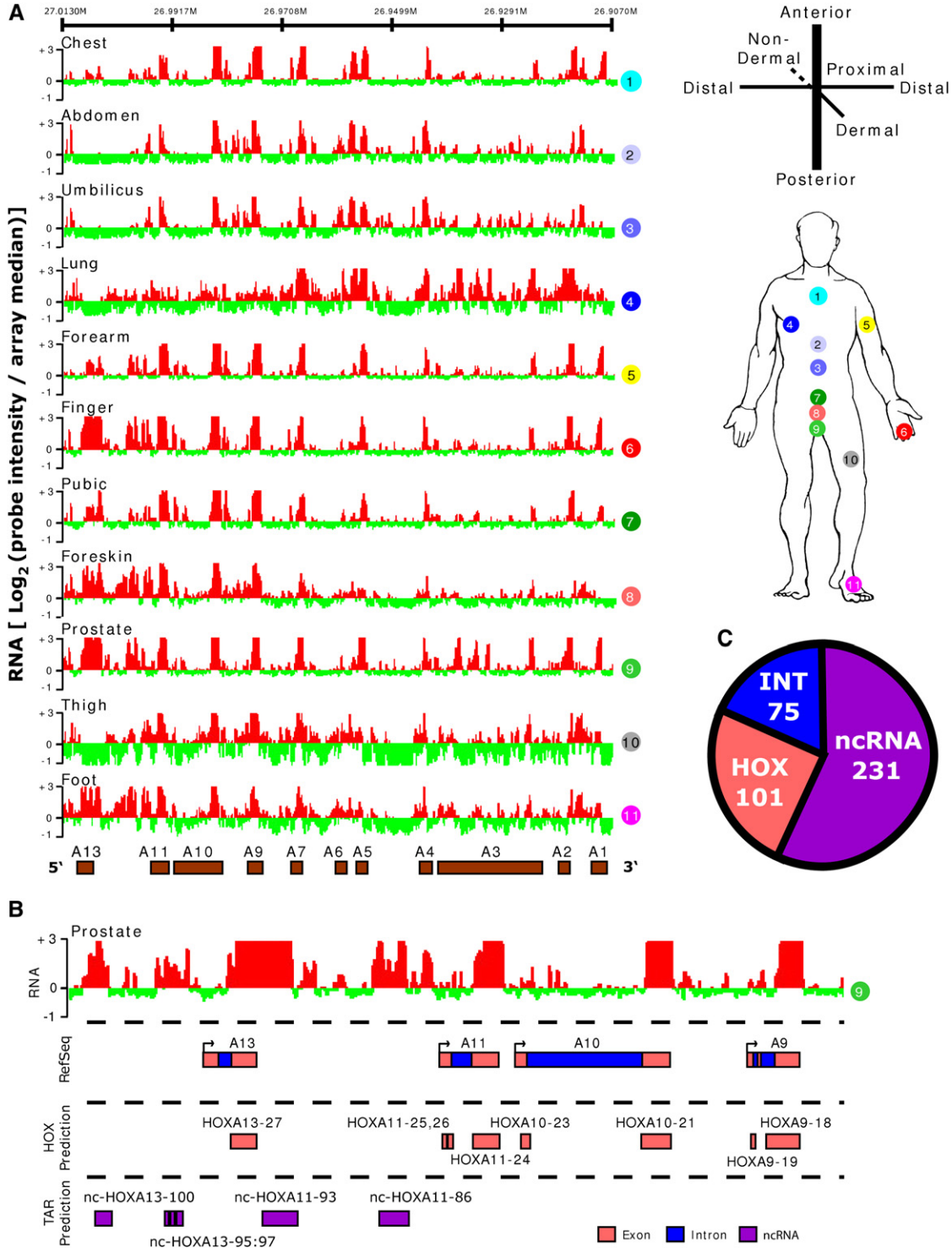

Figure 1. The Human *HOX* Transcriptome

(A) Site-specific transcription of the *HOXA* locus. Left: The hybridization intensity of 50,532 probes that tile the human *HOXA* locus for each of the 11 samples (numbered in circles). The intensity of each probe is displayed as the $\log_2$ of the ratio of the individual probe intensity divided by the average intensity of all 301,027 probes on the array. The $\log_2$ ratio of each probe was averaged over a 100 bp window; red and green bars indicate expression above or below the array mean, respectively. Genomic locations of protein-coding *HOX* genes are displayed as brown boxes. Right: Anatomic origins of the 11 fibroblast samples with respect to the developmental axes.

complexes, the trithorax group (TrxG) of H3K4 HMTase and the PcG H3K27 HMTase, maintain open and closed chromatin domains in the *HOX* loci, respectively, over successive cell divisions (Ringrose and Paro, 2007). Transcription of many ncRNAs has been observed in fly, mouse, and human *HOX* loci (Bae et al., 2002; Bernstein et al., 2005; Carninci et al., 2005; Drewell et al., 2002; Sessa et al., 2006), and three models have been proposed to account for their action based on experiments in *Drosophila*. First, elegant genetic studies suggested that transcription of ncRNAs altered the accessibility of DNA sequences important for TrxG and PcG binding; the act of intergenic transcription enabled TrxG activation of downstream *HOX* genes and prevented PcG-mediated silencing (Ringrose and Paro, 2007; Schmitt et al., 2005). Second, the above model has been extended by a recent report that several ncRNAs transcribed 5′ of the *Drosophila HOX* gene *Ubx* bind to and recruit the TrxG protein Ash1 to the *Ubx* promoter, thereby inducing *Ubx* transcription (Sanchez-Elsner et al., 2006). However, these results have been challenged by the third model of "transcriptional interference," where transcription of 5′ ncRNAs into the promoters of downstream *HOX* genes prevents *HOX* gene expression, leading to transcriptional silencing in *cis* (Petruk et al., 2006). The extent to which any of these models and alternative mechanisms explain the copious amount of transcription in mammalian *HOX* loci remains to be discovered. Nonetheless, the large number of *HOX* ncRNAs, their complex clustering on the chromosomes, and their potentially diverse modes of action suggest ncRNAs play a significant role in *HOX* regulation. By profiling the entire transcriptional and epigenetic landscapes of ~500 kilobase *HOX* loci at near nucleotide resolution, we will begin to discern competing models of ncRNA action in humans and reveal potentially new mechanisms of ncRNA function.

Transcriptomic and proteomic analysis of the *HOX* loci require pure cell populations with distinct positional identities. Rather than study whole animals where cells of many histologic types and positional identities are intermixed, we and others have observed that primary adult human fibroblasts retain many features of the embryonic pattern of *HOX* gene expression both in vitro and in vivo (Bernstein et al., 2005; Chang et al., 2002; Rinn et al., 2006). Differential and colinear expression of *HOX* genes in adult fibroblasts faithfully reflects their position along the anterior-posterior and proximal-distal axes of the developing body (Rinn et al., 2006) and is believed to be important for maintenance of regional identities of skin throughout the lifetime of the animal (Chuong, 2003). The remarkable persistence—over decades—of the embryonic patterns of *HOX* gene expression in these human cells suggests the action of a powerful epigenetic machinery operative over the *HOX* loci. In this study, we create an ultrahigh-resolution tiling microarray to interrogate the transcriptional and epigenetic landscape of the *HOX* loci in a unique collection of primary human fibroblasts with 11 distinct positional identities. Our results identify numerous novel human *HOX* ncRNAs, clarify potential mechanisms of their regulation, and reveal a novel mechanism of ncRNA-assisted transcriptional silencing via the PcG proteins in *trans*.

RESULTS

Noncoding RNAs of the Human *HOX* Loci: Identity, Conservation, Expression Pattern, and Sequence Motifs

To systematically investigate the transcriptional activity of the human *HOX* loci, we designed a DNA microarray for all four human *HOX* loci at five base pair (bp) resolution along with two megabases of control regions (Table S1). Computational and experimental analysis confirmed the specificity of the tiling array to distinguish highly related *HOX* sequences (Figures S1 and S2).

Because adult primary fibroblasts are differentiated based on their anatomic site of origin and retain canonical features of the embryonic *HOX* code (Chang et al., 2002; Rinn et al., 2006), we used *HOX* tiling arrays to profile polyadenylated transcripts from fibroblasts representing 11 distinct positional identities (Figure 1A). Previously, analytic methods for tiling arrays have allowed present/absent calls of transcripts and binding events but were less successful in quantification of signal intensity (Bernstein et al., 2005; Bertone et al., 2004). We addressed this challenge by adapting a signal processing algorithm used in computer vision termed Otsu's method (Otsu, 1979). The method dynamically searches for statistically significant cutoffs between signal and background and detects contiguous regions of at least 100 bp (20 probes) with signal intensity significantly above background. Averaging the signal intensity over all probes in the called region thus produces a quantitative measure of transcript abundance. Using this algorithm, we identified a total of 407 discrete transcribed regions in the four *HOX* loci (Table S2). We used current genome annotations to partition them into known *HOX* gene exons, introns, and intergenic transcripts (Figure 1B). As expected, we detected many transcribed regions that corresponded to known *HOX* exons and introns (101 and 75, respectively), including exonic transcription for 34 of the 39 *HOX* genes, thus indicating that these 11 samples encompass the majority of *HOX* transcriptional activity. In all cases examined, the expression of *HOX* genes as determined by the tiling array matched that previously determined by cDNA microarray

(B) Transcribed regions were identified by contiguous signals on tiling array, then compared with Refseq sequence to define genic (exonic, pink color, and intronic, blue) and intergenic transcribed regions (purple). Each predicted *HOX* exon or intron was named HOXn or int-HOXn, respectively. Intergenic transcribed regions were named as nc-HOXn, where n is the *HOX* paralog located 3′ to the ncRNA on the *HOX*-coding strand.
(C) Summary of transcribed regions in all four *HOX* loci defining the number of *HOX* genic, intronic, and ncRNA-transcribed regions.

and RT-PCR for these same samples (Chang et al., 2002; Rinn et al., 2006).

Interestingly, the majority of the transcribed regions (231 of 407) arise from intergenic regions (Figure 1C). By comparison to databases of all known amino acid sequences, we found that only 13% (29 of 231) of these intergenic transcripts showed any coding potential in all six possible translational frames (Experimental Procedures; Table S3). In contrast, 88% (84 of 96) of the *HOX* exon transcripts had coding potential, where the 12% noncoding exonic transcripts corresponded to untranslated exonic regions. While these results do not completely rule out the possibility of new protein-coding genes interspersed throughout the *HOX* loci, these intergenic transcripts are more likely candidate noncoding RNAs. We therefore refer to these intergenic transcripts as *HOX* ncRNAs. We named each ncRNA by its genomic location, affixing the name of the *HOX* gene located 3′ to the ncRNA on the *HOX*-coding strand (Figure 1C). As previously suggested (Sessa et al., 2006), the majority of ncRNAs (74%) demonstrate evidence for opposite-strand transcription from the *HOX* genes (Table S4). Fifteen percent of the ncRNAs we identified are novel, while the majority of ncRNAs (85%) have been independently observed by EST sequencing or other means (Experimental Procedures). Even for the known ncRNAs, our data suggest that almost all ncRNAs are longer than previously believed (Table S4). The average extension for previously observed ncRNAs is 202 bases; in total we discovered over 30 kilobases of new transcribed bases in the human *HOX* loci. Thus, in just 11 hybridizations, we have substantially expanded the number and length of known transcribed regions in the human *HOX* loci as well as defined their expression patterns throughout the human body.

We found several lines of evidence that confirm the biological importance of the *HOX* ncRNAs. First, comparative analysis with seven vertebrate genomes revealed that some ncRNAs are preferentially conserved in evolution over nontranscribed or intronic *HOX* sequences. For instance, more than one-third of the top 100 conserved transcribed regions in the *HOX* loci are ncRNAs (Figure S3A). Second, RT-PCR analysis of 40 predictions of ncRNA expression levels from the tiling array confirmed a high level of agreement (85%) between array signal intensity and transcript abundance as measured by RT-PCR (Figure S3B).

Third, we found that, like canonical *HOX* genes, ncRNAs also systematically vary their expression along developmental axes of the body in a manner coordinated with their physical location on the chromosome (Figure 2; Table S5). One hundred forty-seven of 231 *HOX* ncRNAs (64%) are differentially expressed along a developmental axis of the body ($p < 0.05$). For instance, 48 *HOX* ncRNAs are differentially expressed with their neighboring *HOX* genes along the proximal-distal axis (close or far from the trunk of the body; $p < 0.05$, Figure 2A). Strikingly, all 41 transcribed regions (both *HOX* genes and ncRNAs) that are induced in distal sites belonged to *HOX* paralogous groups 9–13, and all 30 transcribed regions that are repressed in distal sites belonged to paralogous groups 1–6, precisely recapitulating the evolutionary origin of the two domains from *Drosophila ultrabithorax* and *antennapedia* complexes, respectively ($p < 10^{-19}$, two-way chi-square test; Carroll, 1995). Similarly, 87 *HOX* ncRNAs are differentially expressed along the anterior-posterior axis (top to bottom of the body), this time with ncRNAs from *HOXC9–13* preferentially induced in posterior sites ($p < 0.05$, Figure 2B). Additionally, we observed 7 *HOX* genes and 12 ncRNAs that are either expressed in dermal (outside the body) or nondermal (inside the body) fibroblasts (Figure S4; Table S5). Systematic comparison of the expression pattern of every ncRNA with its immediate 5′ and 3′ *HOX* gene neighbor showed that the vast majority of ncRNAs (90%) are coordinately induced with their 3′ *HOX* genes, while only 10% of instances are ncRNA expression anticorrelated with 3′ *HOX* gene expression (Figure S5).

Fourth, in addition to their distinctive expression patterns, we found that the ncRNAs also possess specific sequence motifs. Using a discriminative motif finder that we previously developed (Segal et al., 2003), we found that ncRNAs are enriched for specific DNA sequence motifs based on their site-specific expression patterns ($p < 10^{-9}$, Figure 3C). We identified a sequence motif enriched in ncRNAs over exonic, intronic, or nontranscribed sequences, and we further identified sequence motifs for ncRNAs that are expressed in distal, proximal, or posterior sites. These sequence motifs may represent DNA- or RNA-binding sites for regulatory factors to regulate gene expression in *cis*. Together, these results establish that the majority of site-specific transcriptional output of the *HOX* loci consists of ncRNAs. Their evolutionary conservation, differential expression along developmental axes, and distinct primary sequence motifs suggest important and possibly widespread roles for these ncRNA transcripts in *HOX* gene regulation.

Diametrical Domains of Chromatin Modifications Demarcated by *HOX* ncRNAs

The coordinate expression of *HOX* genes and neighboring ncRNAs raised the possibility that their expression may be regulated by chromatin domains, large contiguous regions of differential chromatin modifications that enable transcriptional accessibility or cause silencing. Such domains, first observed by Bernstein and colleagues for histone H3 lysine 4 dimethylation (H3K4me2; Bernstein et al., 2005), are a notable and unique feature of *HOX* loci chromatin (Bracken et al., 2006; Lee et al., 2006; Papp and Muller, 2006; Squazzo et al., 2006). We tested this idea by loci-wide chromatin immunoprecipitation followed by tiling array analysis (ChIP-chip). We found that both *HOX* and ncRNA transcription fell within broad domains occupied by RNA polymerase II, whereas the transcriptionally silent regions were broadly occupied by the PRC2 component Suz12 and its cognate histone mark, histone H3 trimethylated at lysine 27 (H3K27me3; $p < 10^{-15}$, chi-square test, Experimental Procedures; Figure 3). Comparison of cells from different anatomic

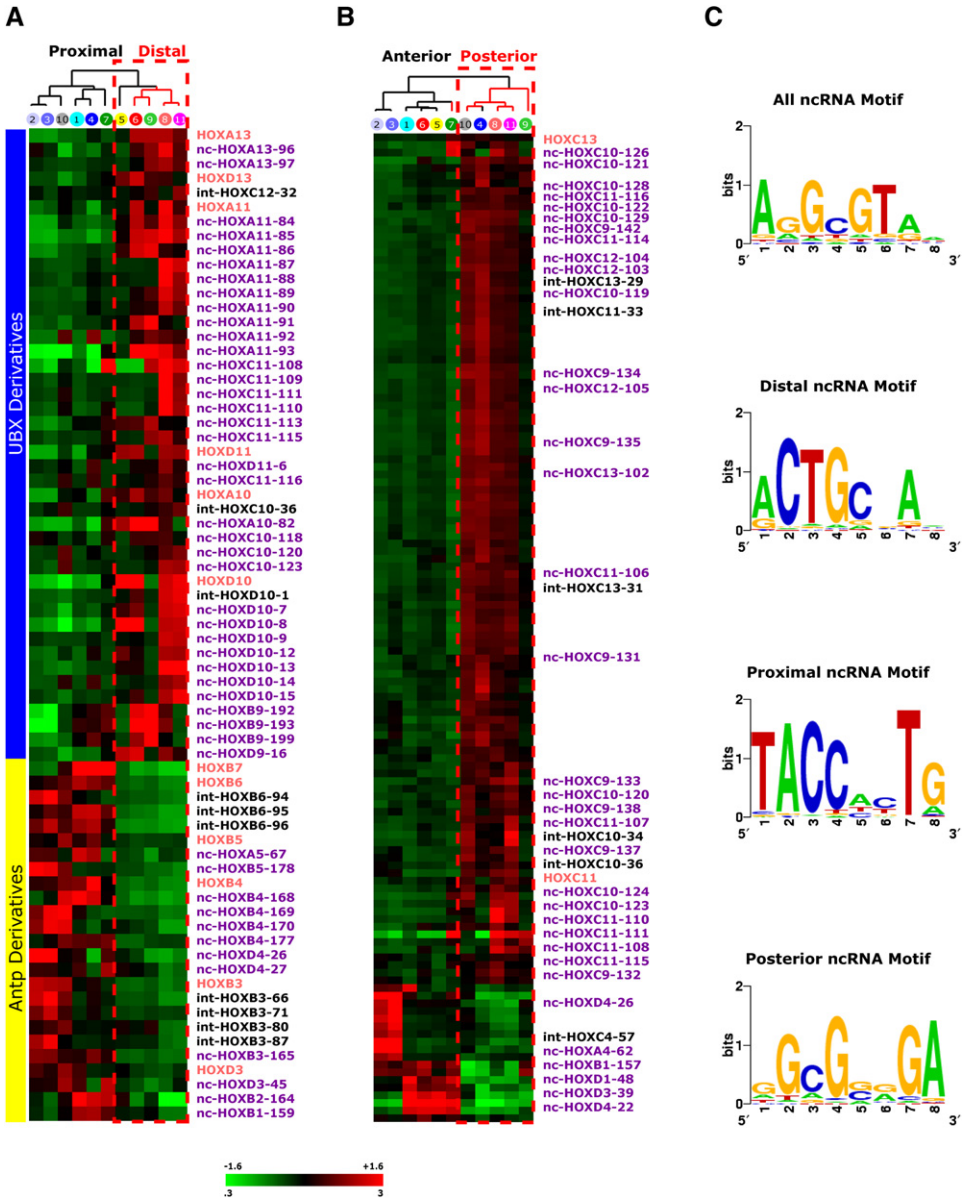

Figure 2. Site-Specific Expression and Primary Sequence Motifs of *HOX* ncRNAs

(A) *HOX*-encoded transcripts differentially expressed along the proximal-distal axis. Sixty transcribed regions (12 *HOX* genes and 48 ncRNAs) were differentially expressed ($p < 0.05$, Student's t test) between distal fibroblast samples (foot, finger, foreskin, and prostate) and all other cells. Expression level of each transcribed region above or below the global median is denoted by the color scale (3-fold to 0.3-fold on linear scale or +1.6 to −1.6 on $\log_2$ scale). Transcribed regions were ordered by their position along the chromosome, and samples were hierarchically clustered by similarity of expression of these 60 transcripts. The evolutionary origin of *HOX* paralogs to fly *ultrabithorax* (UBX) or *antennapedia* (Antp) are indicated by blue and yellow boxes, respectively.

(B) *HOX*-encoded transcripts differentially expressed along the anterior-posterior anatomic division. A total of 92 transcripts (6 *HOX* genes, 86 ncRNAs) were differentially expressed ($p < 0.05$, Student's t test) in anterior or posterior primary fibroblast cultures (above or below the umbilicus). Expression of each ncRNA is represented as in (A).

(C) Enriched sequence motifs in *HOX* ncRNA based on their pattern of expression ($p < 10^{-9}$). Logograms of sequence motifs enriched in the primary sequences of ncRNAs over nontranscribed *HOX* sequences or in ncRNAs with distal, proximial, or posterior patterns of expression are shown. ncRNAs expressed in anterior anatomic sites did not share a primary sequence motif more than expected by chance.

origins showed that the primary DNA sequence can be programmed with precisely the same boundary but in the opposite configuration. For example, in lung fibroblasts the 5′ *HOXA* locus is occupied by Suz12 but not PolII, whereas in foreskin fibroblasts this exact same chromatin domain is occupied by PolII but not Suz12. Thus,

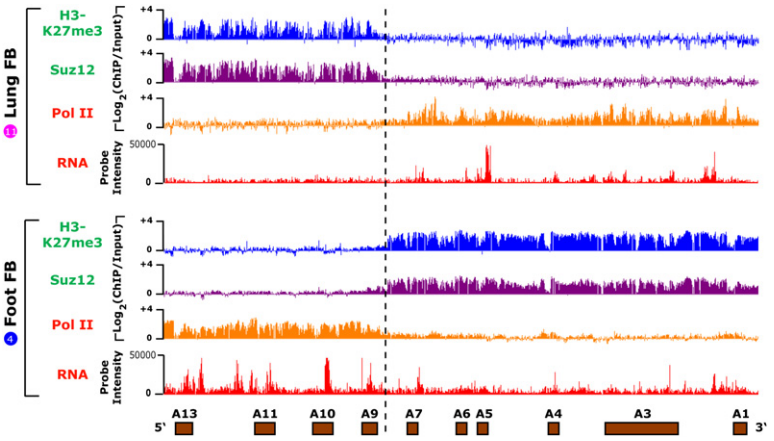

Figure 3. Diametrically Opposed Chromatin Modifications and Transcriptional Accessibility in the *HOXA* Locus

Occupancy of Suz12, H3K27me3, and PolII versus transcriptional activity over ~100 kb of the *HOXA* locus for primary lung (top) or foot (bottom) fibroblasts (Fb). For chIP data, the $\log_2$ ratio of ChIP/Input is plotted on the y axis. For RNA data, the hybridization intensity on a linear scale is shown. Dashed line highlights the boundary of opposite configurations of chromatin modifications and intergenic transcription.

positional identity in differentiated cells may be marked by diametric or mutually exclusive domains of chromatin modifications, which switch their configurations around a center of inversion in a site-specific manner.

Interestingly, the boundary of the diametric chromatin domains defined by ChIP-chip is precisely the same as that suggested by our transcriptional analysis. In the *HOXA* locus, the chromatin boundary and switch between proximal versus distal expression patterns occurs between *HOXA7* and *HOXA9*. Additional ChIP-chip analysis showed the domain of PolII occupancy precisely overlaps the domain of H3K4 dimethylation, but H3K9 trimethylation, a histone modification characteristic of constitutive heterochromatin, is not present on any *HOX* loci in these cells (Figure S6). These results suggest that *HOX* loci transcription in adult fibroblasts is governed by opposing epigenetic modifications over large chromosomal regions and further define the locations of specific boundary elements that delimit chromatin domains.

HOTAIR: A ncRNA that Regulates Chromatin Silencing In *trans*

We next asked whether the coordinate transcription of *HOX* ncRNAs is merely a consequence of the broad chromatin domains or whether the ncRNAs are actively involved in establishing such domains. To address this question, we analyzed in depth the function of a long ncRNA situated at the boundary of two diametrical chromatin domains in the *HOXC* locus (Figure 4A). This ncRNA is transcribed in an antisense manner with respect to the canonical *HOXC* genes; we therefore named it HOTAIR for *HOX Antisense Intergenic RNA*. Molecular cloning and northern blot analysis confirmed that HOTAIR is a 2158 nucleotide, spliced, and polyadenylated transcript; strand-specific RT-PCR analysis confirmed that only one strand of HOTAIR that is antisense to *HOXC* genes is transcribed (Figures 4B and 4C). Computational analysis of HOTAIR secondary structure did not reveal obvious stem loops suggestive of pre-miRNAs. Northern blot analysis of size-fractionated RNA showed no evidence of small RNA products suggestive of micro- or siRNA production, while we readily detected the ubiquitous miRNA let7 in parallel experiments (Figure 4D).

Our tiling array data suggested that HOTAIR is preferentially expressed in posterior and distal sites, and indeed this expression pattern is confirmed by additional RT-PCR experiments (Figure 4E). In situ hybridization of developing mouse embryos confirmed that HOTAIR is expressed in posterior and distal sites, indicating the conservation of anatomic expression pattern from development to adulthood (Figure 4F). Interestingly, this transcript has very high nucleotide conservation in vertebrates (99.5%, 95%, 90%, and 85% sequence identity in chimp, macaque, mouse, and dog genomes, respectively), yet is riddled with stop codons with little amino acid sequence conservation amongst vertebrates (Supplemental Experimental Procedures). These results suggest that HOTAIR may function as a long ncRNA.

HOTAIR ncRNA may regulate gene expression in *HOX* loci in *cis* or *trans*; alternatively, it may be the act of antisense transcription in the *HOXC* locus rather than the ncRNA itself that has a functional role in gene regulation. To distinguish between these possibilities, we depleted HOTAIR ncRNA by RNA interference in primary human fibroblasts and determined the consequences on the transcriptional landscape of the *HOX* loci. Strikingly, while siRNA-mediated depletion of HOTAIR had little effect on transcription of the *HOXC* locus on chromosome 12 compared to wild-type and control siRNA targeting GFP, depletion of HOTAIR led to dramatic transcriptional activation of the *HOXD* locus on chromosome 2 spanning over 40 kb, including *HOXD8*, *HOXD9*, *HOXD10*, *HOXD11,* and multiple ncRNAs (Figures 5A, 5B, and S7). To ensure that this was not an off-target effect of RNA interference, we employed four independent siRNA sequences targeting HOTAIR. Each siRNA depleted HOTAIR ncRNA and led to concomitant *HOXD10* activation as determined by quantitative RT-PCR (Figures 5C and 5D). These observations indicate that HOTAIR ncRNA is required to maintain a transcriptionally silent chromosomal domain in *trans* on the *HOXD* locus.

HOTAIR ncRNA Enhances PRC2 Activity at the *HOXD* Locus

To investigate the molecular mechanisms involved in the HOTAIR-dependent silencing of the *HOXD* locus, we used chromatin immunoprecipitation to interrogate changes to the *HOXD* chromatin structure upon depletion of HOTAIR. Our previous ChIP-chip experiments indicated that in primary foreskin fibroblasts, the entire *HOXD* locus was occupied by both Suz12 and H3K27me3. Depletion of HOTAIR followed by ChIP-chip revealed substantial and global loss of H3K27Me3 occupancy over the *HOXD* locus, with the greatest loss residing in the intergenic region between *HOXD4* and *HOXD8* (Figure 6A). HOTAIR depletion also led to a modest but consistent loss of Suz12 occupancy of the *HOXD* locus (Figures 6B and S8). Importantly, occupancy of H3K27me3 and Suz12 across the silent *HOXB* locus was not affected by HOTAIR depletion in these cells. These results suggest that HOTAIR is selectively required to target PRC2 occupancy and activity to silence transcription of the *HOXD* locus.

Because PcG protein binding to chromatin can involve RNA and HOTAIR ncRNA is required for PRC2 function (i.e., H3K27 trimethylation), we reasoned that HOTAIR may bind to PRC2 and directly regulate Polycomb function. Indeed, native immunoprecipitation of Suz12 from nuclear extracts of two types of primary fibroblasts retrieved associated endogenous HOTAIR ncRNA as detected by RT-PCR, but not nonspecific U1 RNA or DNA (Figure 7A). HOTAIR ncRNA was not retrieved by immunoprecipitation of YY1, which has been suggested to be a component of PRC1 (Sparmann and van Lohuizen, 2006). Suz12 also did not associate with the neighboring *HOXC10* mRNA, indicating that PRC2 binds selectively to *HOXC*-derived transcripts (Figure S9). In the reciprocal experiment, we prepared purified biotinylated sense or antisense HOTAIR RNA by in vitro transcription and probed nuclear extracts of HeLa cells to identify HOTAIR-binding proteins. HOTAIR ncRNA retrieved PRC2 components Suz12 and EZH2 but not YY1 (Figure 7B). Antisense HOTAIR RNA did not retrieve any of the above proteins, indicating that the binding conditions are highly specific. Collectively, these experiments indicate that HOTAIR is physically associated with PRC2 either directly or indirectly; loss of this interaction may reduce the ability of PRC2 to methylate histone tails and silence transcription at the *HOXD* locus.

DISCUSSION

Panoramic Views of the *HOX* Loci by Ultrahigh-Resolution Tiling Arrays

By analyzing the transcriptional and epigenetic landscape of the *HOX* loci at high resolution in cells with many distinct positional identities, we were afforded a panoramic view of multiple layers of regulation involved in maintenance of site-specific gene expression. The *HOX* loci are demarcated by broad chromosomal domains of transcriptional accessibility, marked by extensive occupancy of RNA polymerase II and H3K4 dimethylation and, in a mutually exclusive fashion, by occupancy of PRC2 and H3K27me3. The active, PolII-occupied chromosomal domains are further punctuated by discrete regions of transcription of protein-coding *HOX* genes and a large number of long ncRNAs. Our results confirm the existence of broad chromosomal domains of histone modifications and the occupancy of HMTases over the *Hox* loci observed by previous investigators (Bernstein et al., 2005; Boyer et al., 2006; Guenther et al., 2005; Lee et al., 2006; Squazzo et al., 2006) and extend on these observation in several important ways.

First, by comparing the epigenetic landscape of cells with distinct positional identities, we showed that the broad chromatin domains can be programmed with precisely the same boundary but with diametrically opposite histone modifications and consequences on gene expression. Our data thus functionally pinpoint the locations of chromatin boundary elements in the *HOX* loci, the existence of some of which has been predicted by genetic experiments (Kmita et al., 2000). One such boundary element appears to reside between *HOXA7* and *HOXA9*. This genomic location is also the switching point in the expression of *HOXA* genes between anatomically proximal versus distal patterns and is the boundary of different ancestral origins of *HOX* genes, raising the possibility that boundary elements are features demarcating the ends of ancient transcribed regions. Second, the ability to monitor 11 different *HOX* transcriptomes in the context of the same cell type conferred the unique ability to characterize changes in ncRNA regulation that reflect their position in the human body. This unbiased analysis identified more than 30 kb of new transcriptional activity, revealed ncRNAs conserved in evolution, mapped their anatomic patterns of expression, and uncovered enriched ncRNA sequence motifs correlated with their expression pattern—insights which could not be gleaned from examination of EST sequences alone (Sessa et al., 2006). Our finding of a long ncRNA that acts in *trans* to repress *HOX* genes in a distant locus is mainly due to the ability afforded by the tiling array to comprehensively examine the consequence of any perturbation over all *HOX* loci. The expansion of a handful of *Hox*-encoded ncRNAs in *Drosophila* to hundreds of ncRNAs in human *HOX* loci suggests increasingly important and diverse roles for these regulatory RNAs.

An important limitation of the tiling array approach is that while we have improved identification of transcribed regions, the data do not address the *connectivity* of these regions. The precise start, end, patterns of splicing, and regions of double-stranded overlap between ncRNAs will need to be addressed by detailed molecular studies in the future.

ncRNA Transcription and *HOX* Gene Expression

Noncoding RNAs are emerging as regulatory molecules in specifying specialized chromatin domains (Bernstein and Allis, 2005), but the prevalence of a different

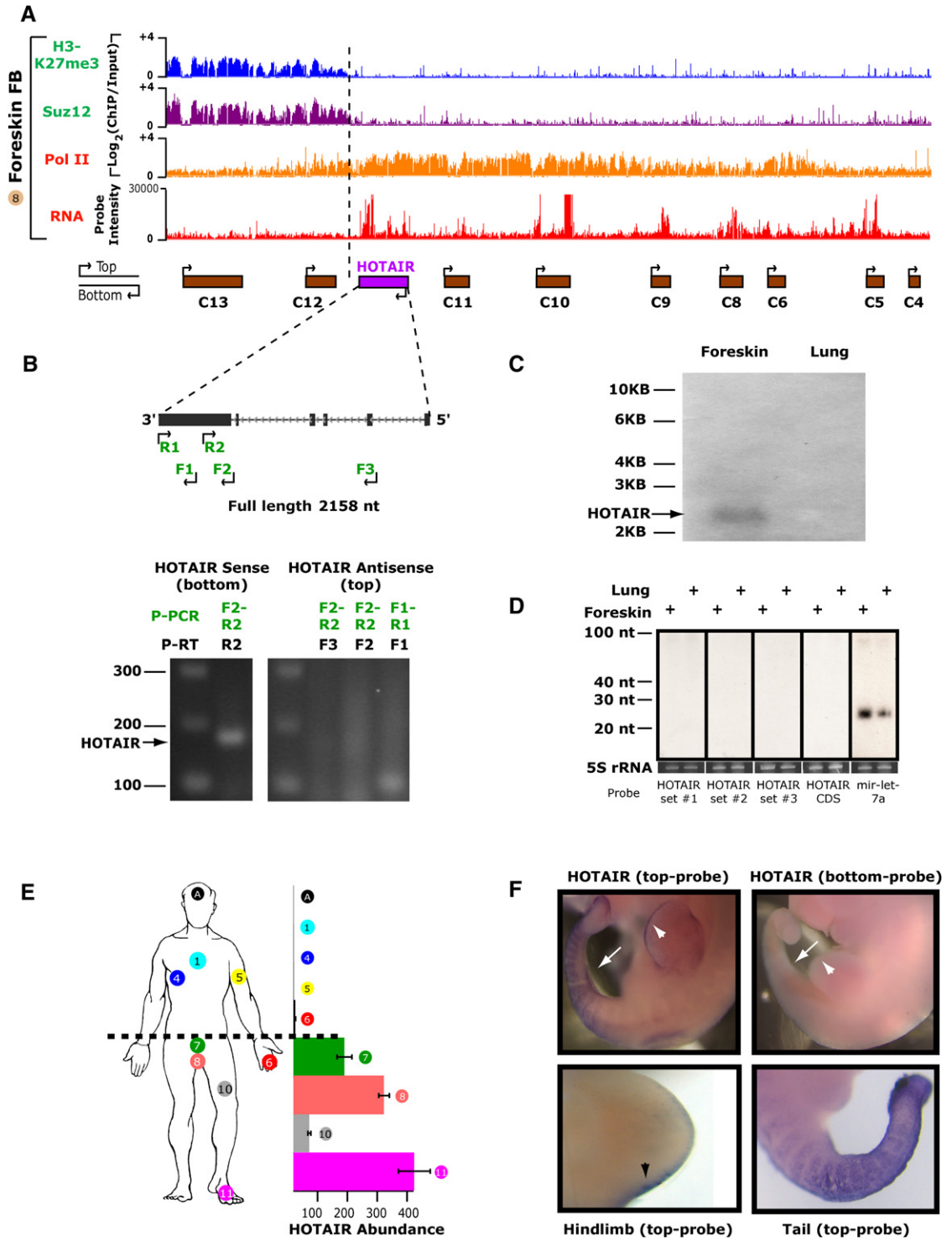

Figure 4. HOTAIR, an Antisense Intergenic Long ncRNA of the *HOXC* Locus

(A) Genomic location of HOTAIR at the boundary of two chromatin domains. ChIP-chip and RNA expression on tiling array are as shown in Figure 3.

(B) Strand-specific RT-PCR shows exclusive expression of HOTAIR from the strand opposite to *HOXC* genes (bottom). Primers for reverse transcription (P-RT) and PCR (P-PCR) were designed to specifically target either the top (primers F1–F3) or bottom strand (primer R1) of HOTAIR.

(C) Northern blot analysis of HOTAIR in lung and foreskin fibroblast RNA.

(D) Size-fractionated small RNA was probed with pools of oligonucleotides spanning HOTAIR (sets #1–3), full-length antisense HOTAIR (CDS), or a probe against miRNA let7a.

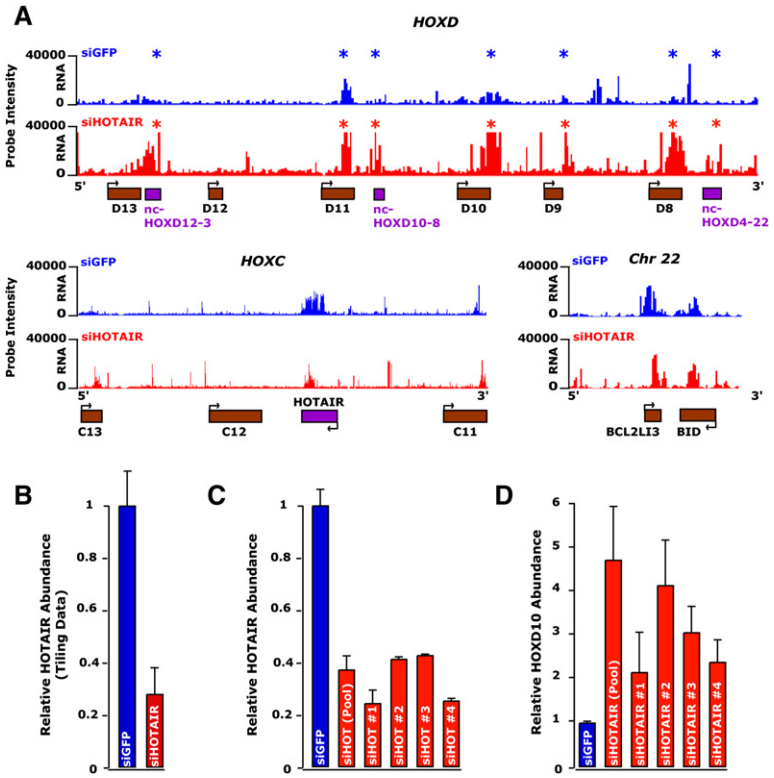

Figure 5. Loss of HOTAIR Results in Transcriptional Induction of *HOXD* Locus

(A) RNA expression profiles of *HOXD* locus (top), *HOXC* locus surrounding HOTAIR (bottom left), and a control region on chromosome 22 (bottom right) following transfection of siRNA targeting GFP (siGFP) or a pool of four siRNAs targeting HOTAIR (siHOTAIR). Intensities of RNA hybridized to the tiling array from the siGFP and the siHOTAIR transfections are plotted on a linear scale in blue and red, respectively. * indicates genes with significant increased transcription.

(B) qRT-PCR measuring the relative abundance of the HOTAIR transcript in the primary foreskin samples shown in (A). Mean ± standard deviation are shown.

(C and D) qRT-PCR measuring the relative abundance of the HOTAIR (C) and *HOXD10* (D) transcripts after depletion of four individual siRNAs to HOTAIR and the pool. Mean ± standard deviation are shown.

mechanism by which they act is not known. In *Drosophila*, transcription of ncRNAs was proposed to induce *HOX* gene expression by activation of *cis*-regulatory elements (Schmitt et al., 2005) or by ncRNA-mediated recruitment of the TrxG protein Ash1 (Sanchez-Elsner et al., 2006). However, an alternative model, termed "transcriptional interference," argues that ncRNA transcription prevents the expression of 3′ located *Hox* genes (Petruk et al., 2006). These two classes of models make opposite predictions on the correlation between expression of 5′ ncRNA and the 3′ *HOX* gene. Our finding of widespread position-specific ncRNAs that flank and are coordinately induced with neighboring human *HOX* genes is consistent with models of *cis* activation by ncRNA transcription. Only 10% of *HOX* ncRNAs demonstrate anticorrelated expression pattern with their cognate 3′ *HOX* genes (Figure S5), suggesting that transcriptional interference is not the main mode of ncRNA action, at least in the cell types that we studied. Our results are also consistent with a recent analysis of *HOX* gene activation during teratocarcinoma cell differentiation, where transcription of certain 5′ ncRNAs immediately preceded *HOX* gene activation (Sessa et al., 2006). Transcriptional interference may be a more prominent mechanism during embryonic development, where its role in *Hox* gene expression was documented in *Drosophila* (Petruk et al., 2006).

Our results uncovered a new mechanism whereby transcription of ncRNA dictates transcriptional silencing of a distant chromosomal domain. The four *HOX* loci demonstrate complex cross regulation and compensation during development (Kmita and Duboule, 2003; Lemons and McGinnis, 2006). For instance, deletion of the entire *HOXC* locus exhibits a milder phenotype than deletion of individual *HOXC* genes, suggesting that there is negative feedback within the locus (Suemori and Noguchi, 2000). Multiple 5′ *HOX* genes, including *HOXC* genes, are expressed in developing limbs (Nelson et al., 1996), and deletion of multiple *HOXA* and *HOXD* genes is required to unveil limb patterning defects (Zakany et al., 1997). Our results suggest that deletion of the 5′ *HOXC* locus, which encompasses HOTAIR, may lead to transcriptional induction of the homologous 5′ *HOXD* genes, thereby restoring the total dosage of HOX transcription factors.

(E) Posterior and distal expression of HOTAIR in human fibroblasts as measured by qRT-PCR. The site of origin of each fibroblast sample is indicated by the sample number on the anatomic cartoon. "A" is derived from the scalp. The relative abundance of HOTAIR in each position, relative to scalp (most anterior), is shown on the x axis.

(F) Whole mount in situ hybridization using HOTAIR sense (bottom strand) or antisense (top strand) probes in embryonic day 10.5 whole-mount embryos (top panels) and the hind limb and tail (bottom left and right panels, respectively). Expression of HOTAIR in posterior hindlimb (arrowhead) and tail (arrow) are highlighted.

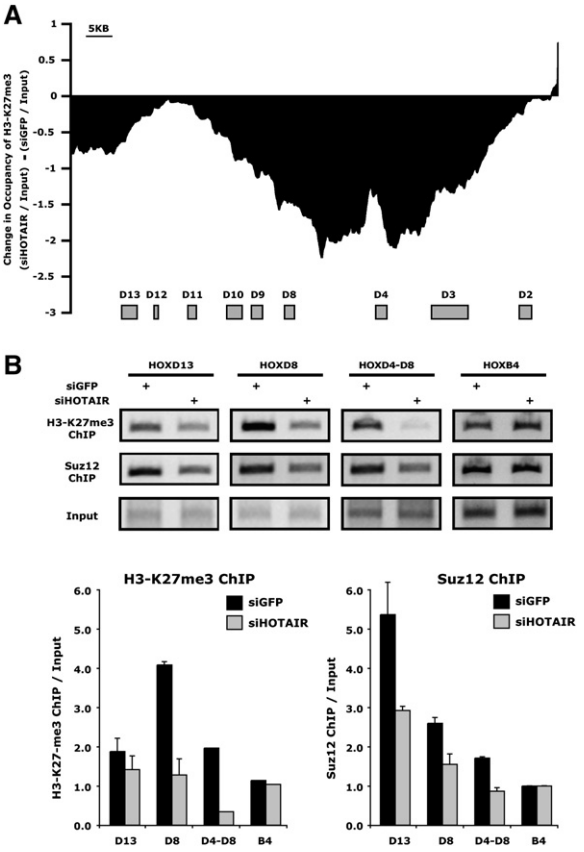

Figure 6. HOTAIR Is Required for H3K27 Trimethylation and Suz12 Occupancy of the *HOXD* Locus

(A) Change in H3K27me3 ChIP-chip signal over the *HOXD* locus caused by depletion of HOTAIR compared to control siRNA against GFP. The location of *HOXD* genes are indicated by boxes.

(B) ChIP of H3K27me3 and Suz12 of select promoters across the *HOXD* locus after siRNA treatment targeting GFP or HOTAIR. Bottom: Quantitation of ChIP assays (mean ± standard error).

How *HOX* ncRNAs may contribute to cross regulation among *HOX* genes should be addressed in future studies.

HOTAIR ncRNA Is Involved in PRC2-Mediated Silencing of Chromatin

Because many HMTase complexes lack DNA-binding domains but possess RNA-binding motifs, it has been postulated that ncRNAs may guide specific histone modification activities to discrete chromatin loci (Bernstein and Allis, 2005; Sun and Zhang, 2005). We have shown that HOTAIR ncRNA binds PRC2 and is required for robust H3K27 trimethylation and transcriptional silencing of the *HOXD* locus. HOTAIR may therefore be one of the long-sought-after RNAs that interface the Polycomb complex with target chromatin. A potentially attractive model of epigenetic control is the programming of active or silencing histone modifications by specific noncoding RNAs (Figure 7C). Just as transcription of certain ncRNA can facilitate H3K4 methylation and activate transcription of the downstream *Hox* genes (Sanchez-Elsner et al., 2006; Schmitt et al., 2005), distant transcription of other ncRNAs may target the H3K27 HMTase PRC2 to specific genomic sites, leading to silencing of transcription and establishment of facultative heterochromatin. In this view, extensive transcription of ncRNAs is both functionally involved in the demarcation of active and silent domains of chromatin as well as a consequence of such chromatin domains.

Several lines of evidence suggest that HOTAIR functions as a bona fide long ncRNA to mediate transcriptional silencing. First, we detected full-length HOTAIR in vivo and in primary cells but not small RNAs derived from HOTAIR indicative of miRNA or siRNA production. Second, depletion of full-length HOTAIR led to loss of *HOXD* silencing and H3K27 trimethyation by PRC2, and third, endogenous or in-vitro-transcribed full-length HOTAIR ncRNA physically associated with PRC2. While these results do not rule out the possibility that RNA interference pathways may be subsequently involved in PcG function (Grimaud et al., 2006; Kim et al., 2006), they support the notion that the long ncRNA form of HOTAIR is functional. The role of HOTAIR is reminiscent of XIST, another long ncRNA shown to be involved in transcriptional silencing of the inactive X chromosome (Plath et al., 2003). An important difference between HOTAIR and XIST is the strictly *cis*-acting nature of XIST. To our knowledge, HOTAIR is the first example of a long ncRNA that can act in *trans* to regulate a chromatin domain. While we have observed a *trans*-repressive role for HOTAIR, our data do not permit us to rule out a *cis*-repressive role in the *HOXC* locus. Our siRNA-mediated depletion of HOTAIR was substantial but incomplete; further, the proximity between the site of HOTAIR transcription and the neighboring *HOXC* locus may ensure significant exposure to HOTAIR even if the total pool of HOTAIR in the cell were depleted. The precise location of HOTAIR at the boundary of a silent chromatin domain in the *HOXC* locus makes a *cis*-repressive role a tantalizing possibility. Judicious gene targeting of HOTAIR may be required to address its role in *cis*-regulation of chromatin.

The discovery of a long ncRNA that can mediate epigenetic silencing of a chromosomal domain in *trans* has several important implications. First, ncRNA guidance of PRC2-mediated epigenetic silencing may operate more globally than just in the *HOX* loci, and it is possible that other ncRNAs may interact with chromatin-modification enzymes to regulate gene expression in *trans*. Second, PcG proteins are important for stem cell pluripotency and cancer development (Sparmann and van Lohuizen, 2006); these PcG activities may also be guided by stem cell- or cancer-specific ncRNAs. Third, Suz12 contains a zinc finger domain, a structural motif that can bind RNA (Hall, 2005), and EZH2 and EED both have in vitro RNA-binding activity (Denisenko et al., 1998). The interaction between HOTAIR and PRC2 may also be indirect and mediated by additional factors. Detailed studies of HOTAIR and PRC2 subunits are required to elucidate the structural features that establish the PRC2 interaction

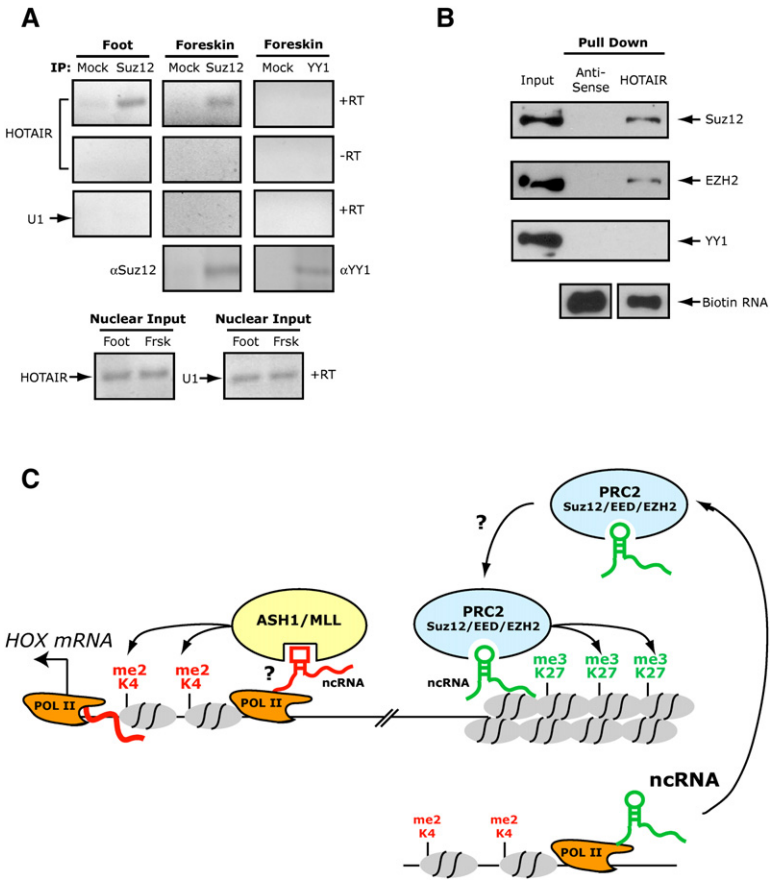

Figure 7. HOTAIR ncRNA Binds PRC2

(A) Immunoprecipitation of Suz12 retrieves endogenous HOTAIR. Nuclear extracts of foot or foreskin fibroblasts were immunoprecipiated by IgG (lanes 1, 3, and 5), anti-Suz12 (lanes 2 and 4), or anti-YY1 (lane 6). Coprecipitated RNAs were detected by RT-PCR using primers for HOTAIR (rows 1 and 2) or U1 small nuclear RNA (row 3). To demonstrate that the HOTAIR band was not due to DNA contamination, each RT-PCR was repeated without reverse transcriptase (−RT, row 2). Immunoprecipitation of Suz12 and YY1 were successful as demonstrated by IP-western using the cognate antibodies (row 4). RT-PCR of nuclear extracts demonstrated equal input RNAs (row 5).

(B) In vitro-transcribed HOTAIR retrieves PRC2 subunits. Immunoblot analysis of the indicated proteins is shown; five percent of input extract (5 μg) was loaded as input control.

(C) Model of long ncRNA regulation of chromatin domains via histone-modification enzymes. Transcription of ncRNAs in *cis* may increase the accessibility of TrxG proteins such as ASH1 or MLL or directly recruit them, leading to H3K4 methylation and transcriptional activation of the downstream *HOX* gene(s). In contrast, recruitment of PRC2 is programmed by ncRNAs produced in *trans*, which targets PRC2 activity by as-yet-incompletely-defined mechanisms to target loci. PRC2 recruitment leads to H3K27 methylation and transcriptional silencing of neighboring *HOX* genes.

with HOTAIR. As we illustrated here, high-throughput approaches for the discovery and characterization of ncRNAs may aid in dissecting the functional roles of ncRNAs in these diverse and important biological processes.

EXPERIMENTAL PROCEDURES

Tiling array design, hybridization, signal processing, RT-PCR validation of ncRNAs, and motif analysis are described in Supplemental Data.

Chromatin Immunoprecipitation

Conventional ChIP and ChIP-chip were performed using anti-H3K27me3 (Upstate Cell Signaling cat# 07-449), anti-Suz12 (Abcam cat# 12,201), anti-PolII (Covance cat# MMS-126R), anti-H3K4me2 (Abcam cat# ab7766), anti-H3K9me3 (Abcam cat# ab1186), and Whole Genome Amplification kit (Sigma) as previously described (Squazzo et al., 2006).

HOTAIR Cloning and Sequence Analysis

5′ and 3′ RACE were performed using the RLM Race kit (Ambion) as recommended by the manufacturer.

HOTAIR Expression Analysis

In situ hybridizations of C57BL/6 mouse embryo using human HOTAIR sequence 164–666 (clone 7T; Albrecht et al., 1997), northern blot using full-length HOTAIR, qRT-PCR with SYBR Green (forward HOTAIR, GGGGCTTCCTTGCTCTTCTTATC; reverse, GGTAGAAAAAGCAAC CACGAAGC), and Taqman analysis of *HOXD10* expression (Applied Biosystems, cat# Hs00157974_m1) were as described (Rinn et al., 2004). Small RNA northern blotting was as described (Lau et al., 2001) with the following modifications: 15 μg of small RNA retained total RNA (mirVana miRNA isolation kit, Ambion) was denatured in Novex sample loading buffer and loaded onto 15% TBE-urea gel in Novex running buffer (Invitrogen). RNA was transferred onto Hybond-XL membrane (Amersham) and probed with pools of ^{32}P-γATP end-labeled 40 mer oligos spanning HOTAIR sequence 1–400 (set 1), 401–800 (set 2), 801–1200 (set 3), full-length HOTAIR probe, or a probe for microRNA let-7a (AACTATACAACCTACT ACCTCA) as positive control.

RNA Interference

Foreskin fibroblasts were transfected with 50 nM of siRNAs targeting HOTAIR (#1 GAACGGGAGUACAGAGAGAUU; #2 CCACAUGAA CGCCCAGAGAUU; #3 UAACAAGACCAGAGAGCUGUU; and #4 GAGGAAAAGGGAAAAUCUAUU) or siGFP (CUACAACAGCCACAAC GUCdTdT) using Dharmafect 3 (Dharmacon, Lafayette, CO, USA) per the manufacturer's direction. Total RNA was harvested for total RNA 72 hr later for microarray analysis as previously described (Rinn et al., 2006).

RNA Immunoprecipitation

Foreskin and foot fibroblasts were grown as previously described (Rinn et al., 2006). 10^7 cells were harvested by trypsinization and resuspended in 2 ml PBS, 2 ml nuclear isolation buffer (1.28 M sucrose; 40 mM Tris-HCl pH 7.5; 20 mM MgCl2; 4% Triton X-100), and 6 ml water on ice for 20 min (with frequent mixing). Nuclei were pelleted by centrifugation at 2,500 G for 15 min. Nuclear pellet was

resuspended in 1 ml RIP buffer (150 mM Kcl, 25 mM Tris pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40, 9 ug/ml leupeptin, 9 ug/ml pepstatin, 10 ug/ml chymostatin, 3 ug/ml aprotinin, 1 mM PMSF, 100 U/ml SUPERASin; Ambion). Resuspended nuclei were split into two fractions of 500 μl each (for Mock and IP) and were mechanically sheared using a dounce homogenizer with 15–20 strokes. Nuclear membrane and debris were pelleted by centrifugation at 13,000 RPM for 10 min. Antibody to Suz12 (Abcam cat# 12,201), YY1 (Santa Cruz Biotechnology cat# sc1703), or FLAG epitope (Mock IP, Sigma) was added to supernatant (Suz12: 6 μg, YY1: 10 μg) and incubated for 2 hr at 4°C with gentle rotation. Forty microliters of protein A/G beads were added and incubated for 1 hr at 4°C with gentle rotation. Beads were pelleted at 2,500 RPM for 30 s, the supernatant was removed, and beads were resuspended in 500 μl RIP buffer and repeated for a total of three RIP washes, followed by one wash in PBS. Beads were resuspended in 1 ml of Trizol. Coprecipitated RNAs were isolated, and RT-PCR for HOTAIR (Forward, GGGGCTTCCTTGCTCTTCTTATC; reverse GGTA GAAAAAGCAACCACGAAGC) or U1 (forward, ATACTTACCTGG CAGGGGAG; reverse, CAGGGGGAAAGCGCGAACGCA) was performed as described (Rinn et al., 2006). Protein isolated by the beads was detected by western blot analysis.

HOTAIR RNA Pull-Down of PcG Proteins

Biotin-labeled, full-length HOTAIR RNA and antisense HOTAIR fragment (clone 7T) were prepared with the Biotin RNA Labeling Mix (Roche) and T7 RNA polymerase (Stratagene). Biotinylated RNAs were treated with RNase-free DNase I and purified on G-50 Sephadex Quick Spin columns (Roche). Ten picomole biotinylated RNA was heated to 60°C for 10 min and slow-cooled to 4°C. RNA was mixed with 100 μg of precleared transcription and splicing-competent HeLa nuclear extract (Gozani et al., 1994) in RIP buffer supplemented with tRNA (0.1 μg/ul) and incubated at 4°C for 1 hr. Sixty microliters washed Streptavidin agarose beads (Invitrogen) were added to each binding reaction and further incubated at 4°C for 1 hr. Beads were washed briefly five times in Handee spin columns (Pierce) and boiled in SDS buffer, and the retrieved protein was visualized by immunoblotting.

URLs

All primary data are available at the Stanford Microarray Database (http://genome-www5.stanford.edu/) and Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). Full-length sequence of human HOTAIR ncRNA has been deposited in Genebank (Accession # DQ926657 (bankit841140)).

Supplemental Data

Supplemental Data include Experimental Procedures, nine figures, five tables, and References and can be found with this article online at http://www.cell.com/cgi/content/full/129/7/1311/DC1/.

ACKNOWLEDGMENTS

We thank A.S. Adler, M.L. Cleary, O. Gozani, D. Herschlag, D. Hogan, P.A. Khavari, A.E. Oro, O.J. Rando, and C. Woo for discussion and critical review of the manuscript. Supported by grants from the National Institutes of Health (J.A.H., P.J.F., E.S., and H.Y.C.), Israel Science Foundation (M.K. and E.S.), and National Science Foundation (J.K.W.). E.S. is the incumbent of the Soretta and Henry Shapiro Career Development Chair; J.L.R. is a Fellow, and H.Y.C. is the Kenneth G. and Elaine A. Langone Scholar of the Damon Runyon Cancer Research Foundation. H.Y.C. is also supported by the Emerald Foundation.

Received: January 19, 2007
Revised: March 28, 2007
Accepted: May 9, 2007
Published: June 28, 2007

REFERENCES

Albrecht, U.E.G., Helms, J.A., and Lin, H. (1997). Visualization of gene expression patterns by in situ hybridization. In Molecular and Cellular Methods in Developmental Toxicology, G.P. Daston, ed. (Boca Raton, FL: CRC Press), pp. 23–48.

Bae, E., Calhoun, V.C., Levine, M., Lewis, E.B., and Drewell, R.A. (2002). Characterization of the intergenic RNA profile at abdominal-A and abdominal-B in the Drosophila bithorax complex. Proc. Natl. Acad. Sci. USA *99*, 16847–16852.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169–181.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev. *19*, 1635–1655.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. *26*, 2560–2569.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. Science *306*, 2242–2246.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Carroll, S.B. (1995). Homeotic genes and the evolution of arthropods and chordates. Nature *376*, 479–485.

Chang, H.Y., Chi, J.T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D., and Brown, P.O. (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. Proc. Natl. Acad. Sci. USA *99*, 12877–12882.

Chuong, C.M. (2003). Homeobox genes, fetal wound healing, and skin regional specificity. J. Invest. Dermatol. *120*, 9–11.

Denisenko, O., Shnyreva, M., Suzuki, H., and Bomsztyk, K. (1998). Point mutations in the WD40 domain of Eed block its interaction with Ezh2. Mol. Cell. Biol. *18*, 5634–5642.

Drewell, R.A., Bae, E., Burr, J., and Lewis, E.B. (2002). Transcription defines the embryonic domains of cis-regulatory activity at the Drosophila bithorax complex. Proc. Natl. Acad. Sci. USA *99*, 16853–16858.

Gozani, O., Patton, J.G., and Reed, R. (1994). A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. EMBO J. *13*, 3356–3367.

Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. Cell *124*, 957–971.

Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E., and Young, R.A. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. Proc. Natl. Acad. Sci. USA *102*, 8603–8608.

Hall, T.M. (2005). Multiple modes of RNA recognition by zinc finger proteins. Curr. Opin. Struct. Biol. *15*, 367–373.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T.R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res. *15*, 987–997.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876–880.

Kim, D.H., Villeneuve, L.M., Morris, K.V., and Rossi, J.J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat. Struct. Mol. Biol. *13*, 793–797.

Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. Science *301*, 331–333.

Kmita, M., Kondo, T., and Duboule, D. (2000). Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. Nat. Genet. *26*, 451–454.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858–862.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Lemons, D., and McGinnis, W. (2006). Genomic evolution of Hox gene clusters. Science *313*, 1918–1922.

Nelson, C.E., Morgan, B.A., Burke, A.C., Laufer, E., DiMambro, E., Murtaugh, L.C., Gonzales, E., Tessarollo, L., Parada, L.F., and Tabin, C. (1996). Analysis of Hox gene expression in the chick limb bud. Development *122*, 1449–1466.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Trans. Sys. Man. Cybern. *9*, 62–66.

Papp, B., and Muller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. *20*, 2041–2054.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell *127*, 1209–1221.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223–232.

Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M., et al. (2003). The transcriptional activity of human Chromosome 22. Genes Dev. *17*, 529–540.

Rinn, J.L., Rozowsky, J.S., Laurenzi, I.J., Petersen, P.H., Zou, K., Zhong, W., Gerstein, M., and Snyder, M. (2004). Major molecular differences between mammalian sexes are involved in drug metabolism and renal function. Dev. Cell *6*, 791–800.

Rinn, J.L., Bondre, C., Gladstone, H.B., Brown, P.O., and Chang, H.Y. (2006). Anatomic demarcation by positional variation in fibroblast gene expression programs. PLoS Genet. *2*, e119. 10.1371/journal.pgen.0020119.

Sanchez-Elsner, T., Gou, D., Kremmer, E., and Sauer, F. (2006). Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. Science *311*, 1118–1123.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.

Segal, E., Yelensky, R., and Koller, D. (2003). Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. Bioinformatics *19* (*Suppl 1*), i273–i282.

Sessa, L., Breiling, A., Lavorgna, G., Silvestri, L., Casari, G., and Orlando, V. (2006). Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. RNA *13*, 223–239.

Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. Nat. Rev. Cancer *6*, 846–856.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.W., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res *16*, 890–900.

Suemori, H., and Noguchi, S. (2000). Hox C cluster genes are dispensable for overall body plan of mouse embryonic development. Dev. Biol. *220*, 333–342.

Sun, Y., and Zhang, H. (2005). A unified mode of epigenetic gene silencing: RNA meets polycomb group proteins. RNA Biol *2*, 8–10.

Zakany, J., Fromental-Ramain, C., Warot, X., and Duboule, D. (1997). Regulation of number and size of digits by posterior Hox genes: a dose-dependent mechanism with potential evolutionary implications. Proc. Natl. Acad. Sci. USA *94*, 13695–13700.

Zhang, H., Christoforou, A., Aravind, L., Emmons, S.W., van den Heuvel, S., and Haber, D.A. (2004). The C. elegans Polycomb gene SOP-2 encodes an RNA binding protein. Mol. Cell *14*, 841–847.

Accession Numbers

Full-length sequence of human HOTAIR ncRNA has been deposited in Genebank (Accession # DQ926657 (bankit841140)).

Montagna Symposium on the Biology of Skin

A Systems Biology Approach to Anatomic Diversity of Skin

John L. Rinn[1], Jordon K. Wang[1], Helen Liu[1], Kelli Montgomery[2], Matt van de Rijn[2] and Howard Y. Chang[1]

Human skin exhibits exquisite site-specific morphologies and functions. How are these site-specific differences specified during development, maintained in adult homeostasis, and potentially perturbed by disease processes? Here, we review progress in understanding the anatomic patterning of fibroblasts, a major constituent cell type of the dermis and key participant in epithelial–mesenchymal interactions. The gene expression programs of human fibroblasts largely reflect the superimposition of three gene expression profiles that demarcate the fibroblast's position relative to three developmental axes. The HOX family of homeodomain transcription factors is implicated in specifying site-specific transcriptional programs. The use of gene, tiling, and tissue microarrays together gives a comprehensive view of the gene regulation involved in patterning the skin.

Journal of Investigative Dermatology (2008) **128,** 776–782; doi:10.1038/sj.jid.5700986

Introduction

The human skin shows remarkable diversity in its structure and function across anatomic sites. Scalp skin is easily recognizable by the numerous terminal hair follicles; in contrast palmoplantar skin possesses no hair follicles but is characterized by increased number of eccrine glands and compact hyperkeratosis of the stratum corneum. The anatomic diversity of human skin is also reflected in the site specificity of many skin diseases and their response to treatment; the distribution of skin lesions is indeed often one of the key clues to the correct diagnosis. From a basic science perspective, the anatomic diversity of skin raises many intriguing questions about how cells acquire and maintain their positional identities in a complex, self-renewing tissue. Here, we review recent progress in understanding site-specific differentiation of cell types in the skin, focusing on emerging systems biology approaches that are beginning to provide comprehensive descriptions of this fascinating biology.

Studies of skin development have shown that site-specific differentiation of epithelia critically depends on epithelial–mesenchymal interactions. Hair and other skin appendages develop through a complex series of reciprocal interactions between epidermal cells and fibroblasts, beginning with a signal from dermal fibroblasts to the overlying epidermis to proliferate, forming the placodes that are progenitors of hair follicles (Millar, 2002). Classic heterotopic recombination

Editor's Note

In 1950, Dr William Montagna, a biologist at Brown University, began a symposium that was designed to bring together basic scientists and clinically trained dermatologists to further our understanding of skin (see Origin of the annual symposium on the biology of skin. Kligman AM (2002) *J Invest Dermatol Symp Proc* 7:1). This month we celebrate the 56th Annual Montagna Symposium on the Biology of Skin, which focused on the development and diseases of skin appendages, with two Perspectives articles. Dr Montagna would be pleased with such a topic, as much of his investigative interest was focused on skin appendages, including hair and sebaceous glands, and on the development of skin, from mice to primates. In the first article, Rinn and co-authors highlight how the emerging science of systems biology can be used to study and understand the wide diversity that is present in skin. In the second article, Wang and co-authors describe how signaling in the skin through Smads can impact skin appendage development. These articles highlight how the application of new technology and new signaling pathways can be applied to skin biology to help us understand the complexity of skin.

Russell P. Hall III,
Deputy Editor

[1]*Program in Epithelial Biology, Department of Dermatology, Stanford University, Stanford, California, USA and* [2]*Program in Epithelial Biology, Department of Pathology, Stanford University, Stanford, California, USA*

Correspondence: Professor Howard Y. Chang, CCSR2155c, 269 Campus Drive, Stanford, California 94305, USA. E-mail: howchang@stanford.edu

Abbreviations: EC, endothelial cell; nt, nucleotide

Received 15 February 2007; revised 10 April 2007; accepted 5 May 2007

experiments showed that a primary dermal signal dictates the overlying epithelial fate; for instance, grafting of wing epidermis to foot dermis in chick led to the development of scales rather than feathers (Dhouailly, 1973, 1984). Specialized fibroblasts from the dermal papilla of hair follicles (Jahoda *et al.*, 1984), but not dermal fibroblasts a few hundred microns away, are able to induce *de novo* hair follicle development when transplanted into naïve skin with epidermal stem cells (Blanpain *et al.*, 2004). Moreover, the type of hair varies throughout the body, and this positional information also is dictated by the local fibroblasts. For example, dermal papilla fibroblasts derived from whiskers induce long, thick whisker-like hairs when transplanted into recipient animals at heterotopic sites (Jahoda, 1992). These intimate and specific epithelial–mesenchymal interactions are not confined only to skin, but appear to be a major theme in the development and homeostasis of all epithelial organs: local fibroblast-like cells in the urogenital sinus induce differentiation of prostatic epithelial precursors to form the prostate gland (Cunha, 1994), and local fibroblasts are also responsible for metanephric induction and pruning of nephrons in the kidney (Schedl and Hastie, 2000; Levinson and Mendelsohn, 2003). Branching morphogenesis of the lung similarly depends critically on reciprocal interactions between bud epithelial cells and surrounding fibroblasts (Shannon and Hyatt, 2004). Because the epidermis is continually shed and replaced by newly developed keratinocytes (every 28 days in humans), it stands to reason that the site-specific inductive capacity of fibroblasts must persist into adulthood, perhaps through the entire lifetime. For instance, cell–cell contact between adult palmoplantar fibroblasts with trunk keratinocytes reprograms these keratinocytes to express palmoplantar keratin genes (Yamaguchi *et al.*, 1999).

Fibroblasts are the principal cell type in the dermis and stroma of internal organs that synthesize connective tissue proteins such as collagen (Dunphy, 1963; Gabbiani and Rungger-Brindle, 1981). Fibroblasts were first described histologically in 1847 by Schwann as ''spindle-shaped or longish corpuscles which are thickest in the middle and gradually elongated in both extremities into minute fibres'' (Schwann, 1847). In practice, fibroblasts are usually identified by their spindle-shaped morphology, the ability to adhere to plastic tissue culture vessels, and the absence of markers for other cell lineages (Dunphy, 1963; Gabbiani and Rungger-Brindle, 1981; Normand and Karasek, 1995). Given these rather non-selective criteria, it is not surprising that cells we currently consider ''fibroblasts'' may comprise a diverse group of cells, each with distinct patterns of synthetic activities and functions.

Site-specific differentiation of fibroblasts

A first test of this idea came with the comprehensive gene expression analysis of primary human dermal fibroblasts from distinct anatomic sites (Chang *et al.*, 2002). Using cDNA array technology, we analyzed the expression of ~21,000 genes in 50 primary human fibroblasts culture from 10 anatomic sites. The results revealed that although all fibroblasts are morphologically similar, the gene expression patterns of cultured fibroblasts from different sites are strikingly distinct. Approximately 8% of all genes transcribed in fibroblasts are differentially expressed in a site-specific manner. The variation and magnitude of gene expression differences among fibroblasts are comparable to the variation observed among different types of white blood cells. When their gene expression patterns were grouped by similarity using a technique called hierarchical clustering (Eisen *et al.*, 1998), fibroblasts from the same topographic sites of the skin were consistently grouped together, and the distinctiveness of topographic gene expression was not obscured among different donors, by passage in tissue culture, or by environmental changes such as serum starvation. In four cases where fibroblasts were derived from multiple sites of the same individual, fibroblasts from the same site of different individuals were far more similar to each other than fibroblasts from different sites of the same individual. These results demonstrate that in fact there are many different cell types that go under the traditional heading of ''fibroblasts''. The main rule of differentiation among dermal fibroblasts appears to be dictated by the anatomic site of origin.

The genes expressed by fibroblasts in a site-specific manner demonstrate distinct choreographed programs in extracellular matrix synthesis, lipid metabolism, and signaling pathways controlling cell migration and cell fate specification (Chang *et al.*, 2002). The gene expression patterns particular to distinct types of fibroblasts were biologically consistent with their anatomical origin. The site-specific expression of many cell growth and differentiation molecules such as members of TGF-β, Wnt, receptor tyrosine kinase and phosphatase families indicates that fibroblasts make important contributions in mesenchymal induction of epithelia. Moreover, the expression domains of genes underlying several genetic diseases affecting skin or musculoskeletal connective tissue correlated closely with the phenotypic defects. For example, by comparing genes that were induced in dermal fibroblasts compared to lung fibroblasts, we were able to identify the genes involved in 6 out of 10 types of Ehlers–Danlos syndrome, a congenital disease characterized by skin fragility and joint laxity. Similarly, we observed that HOXA13 is induced in toe and foreskin fibroblasts, and mutation of HOXA13 in humans leads to hand–foot–genital syndrome, a disease characterized by syndactyly, hypospadias, and malformations of the urogenital system. These results indicate that many important aspects of site-specific differentiation in fibroblasts are preserved *in vitro* and thus amenable to dissection by molecular and functional genomic approaches.

Many epithelial–mesenchymal interactions that specify epidermal appendages occur within local signaling environments of even finer specialization of fibroblasts. Overlaid on the regional fibroblast specialization, recent genomic profiling experiments have indeed identified gene expression signatures of such specialized cells, such as that of the

dermal papilla cells (Rendl *et al.*, 2005). Similarly, we and others have identified unique gene expression signatures of stromal fibroblasts of basal-cell carcinomas, a cancer of hair follicle origin, that distinguished such cells from local fibroblasts of the face (Sneddon *et al.*, 2006).

In contrast, genome-wide expression profiling of purified endothelial cells (ECs) revealed that unlike fibroblasts, EC differentiation is primarily dictated by the vessel size of origin. ECs derived from large vessels from various anatomic locations are more similar to each other than ECs from microvasculature. In particular, dermal microvascular ECs from different anatomic sites were not particularly distinct and were similar to microvascular ECs from intestine and lung (Chi *et al.*, 2003). These results reveal the contrasting regulatory logic of different stromal cell types and reinforce the uniqueness and likely developmental importance of the site-specific differentiation of fibroblasts.

Patterning of adult fibroblasts along three developmental axes

In his influential treatise on development Lewis Wolpert (1969), famously postulated two potential mechanisms of achieving pattern formation: spatial organization of cellular differentiation may be achieved by unique specification of each cell type; alternatively, organization may arise by cells interpreting their position relative to reference points, and adopting specific differentiation programs based on their positional identity within a coordinate system. A particularly attractive and distinctive feature of the positional identity model is the parsimonious use of molecular entities to construct the system, leading to universality of the coordinate system. For example, the same reference points can be used to specify the proximal–distal axis of the upper and lower limbs even though the limbs are spatially distant from each other. Similarly, the same reference points can pattern the anterior–posterior axis of the trunk and also distinguish the upper and the lower limbs.

We hypothesized that these models of pattern formation can be distinguished by comparing the global gene expression profiles of cells distributed throughout the body. Pattern formation by unique specification predicts that the similarity of gene expression profiles of cells would be sporadic and not related by distance from their sites of origin. Conversely, if the positional model of specification is at work, expression profiles will exhibit spatial relationships between fibroblasts in similar positional quadrants of the body.

To distinguish between these models, we analyzed the genome-wide gene expression profiles of 47 primary fibroblast cultures from 43 unique anatomical sites spanning the entire human body (Figure 1a) (Rinn *et al.*, 2006). We confirmed at a much finer level that fibroblasts exhibited large-scale differences in their gene expression programs in a site-specific fashion. Moreover, comparison of gene expression differences among cells from local vicinities *versus* those far away revealed that these gene expression differences are globally related to three recurring segmental patterns: proximal–distal, anterior–posterior (rostral–caudal), and internal–external (non-dermal *vs* dermal) (Figure 1b). For instance, fibroblasts originating from distal upper limb (hand, fingers) had a distinct gene expression signature compared to those from more proximal sites of the arm; fibroblasts from the leg also showed distinct gene expression signatures that portioned them to those above and below the ankle. Remarkably, comparison of the distal arm and distal leg gene signatures showed that these signatures were largely the same, suggesting that this was a signature that reflected distal position along a developmental axis, regardless of the actual position on upper or lower limb. Similarly, the gene expression data that automatically distinguished fibroblasts originating from the top half of the body (anterior or rostral) had a distinctive gene expression profile than from those from the bottom half of the body (posterior or caudal). A third gene expression signature distinguished dermal fibroblasts from fibroblasts that originate from internal organs. These findings suggest

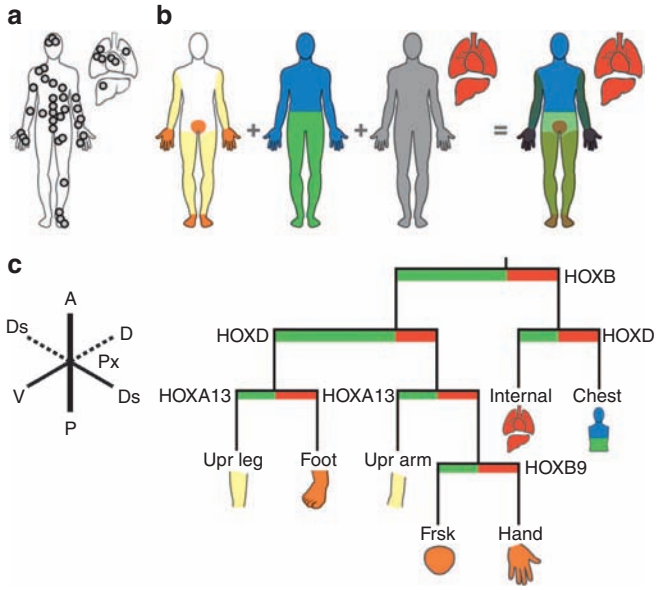

Figure 1. The embryonic pattern of HOX gene expression is retained in adult human fibroblasts. (**a**) A total of 47 primary cultures representing 43 unique anatomic sites (gray circles) that finely map the human body were profiled by gene expression microarrays. (**b**) Model of fibroblast differentiation by overlapping positional patterns of gene expression: proximal (yellow), distal (orange), anterior (blue), posterior (green) and internal organs (red). (**c**) A decision tree of HOX expression that distinguish unique anatomic positions. Right: red indicates higher-than-average and green lower-than-average expression of a given HOX gene, relative to the average expression level across 47 cultures. Each anatomic site can be correctly identified by monitoring the expression level of three HOX genes or less. Left: the developmental axes that demarcate site-specific gene expression of fibroblasts: **a**nterior–**p**osterior, **p**roximal–**d**istal and dermal–non-dermal. A third developmental axis, **d**orsal–**v**entral, did not correlate with large-scale site-specific gene expression in fibroblasts.

that site-specific variations in fibroblast gene expression programs are not idiosyncratic, but rather are systematically related to their positional identities relative to major anatomic axes. Much like a global positioning system, the site-specific gene expression program of a fibroblast reflects the superimposition of three positional coordinates, potentially providing critical cues for development, trafficking, and homeostasis of surrounding cells in the skin.

The positional identities of adult fibroblasts raise the question of whether their cognate coordinate system was established during embryonic development. During embryogenesis, expression of specific HOX genes demarcates distinct positional identities that lead to site-specific cellular differentiation and tissue morphogenesis. The HOX family of homeodomain transcription factors comprise 39 genes that are clustered on four chromosomal loci; their physical order on the chromosomes reflects their spatial pattern of expression along the anterior–posterior and proximal–distal axes of the embryo—a property called colinearity. We found that a very similar pattern of HOX gene expression was also retained in fibroblasts both *in vitro* and *in vivo* (Rinn *et al.*, 2006). In fact, HOX gene expression alone was predictive of the anatomic origin of a given fibroblast culture (Figure 1c). For example, both dermal and non-dermal fibroblasts from the trunk express HOXB genes. However, HOXD gene expression is limited to dermal fibroblasts and is not found in non-dermal fibroblasts. Moreover, distal fibroblasts can be distinguished from proximal fibroblasts by HOXA13 gene expression. Thus, using a very simple decision tree of HOX gene expression, the anatomic position of a fibroblast can be predicted (Figure 1c). HOX genes may also be globally expressed in skin in select circumstances, such as the HOXC13 gene during terminal hair follicle development (Godwin and Capecchi, 1998); however, we mainly observed canonical, site-specific expression of HOX genes in fibroblasts. The salient description of fibroblast anatomic origin by HOX gene expression suggests that this family of transcription factors may have a role in the establishment of site-specific fibroblast gene expression programs.

Revealing the epigenetic landscape by tiling microarrays

Major features of the embryonic pattern of HOX gene expression are retained in adult human fibroblasts from young and old donors alike, and these site-specific patterns of HOX gene expression are also preserved after extensive *ex vivo* cell divisions (J.L.R. and H.Y.C., unpublished observation). The stability of the site-specific transcriptional patterns suggests the possibility of epigenetic mechanisms in their maintenance. Classic genetic studies in *Drosophila* established that HOX gene expression patterns are maintained by two opposing histone modification activities. The Polycomb group proteins mediate histone H3 lysine 27 methylation and are required for the maintenance of HOX gene repression, whereas the Trithorax proteins mediate histone H3 lysine 4 methylation and are required for maintenance of HOX gene activation (Ringrose and Paro, 2004). In addition, intergenic transcription of long noncoding RNAs also plays important roles in the epigenetic maintenance of HOX transcriptional patterns (Schmitt and Paro, 2006). While some of these components are known in model organisms, the transcriptional and epigenetic landscapes in human cells are much less understood. A comprehensive view of transcription and chromatin structure of the HOX loci is needed to decipher the regulatory mechanisms that result in the spatial and temporal patterning of HOX gene expression.

Primary cells from distinct anatomic sites of human skin offer a unique resource of purified cells of specific positional identities that may be used to interrogate the epigenetic mechanisms of site-specific HOX expression programs. We and others have used a technology called tiling microarrays to gain a comprehensive view of the HOX loci (Bernstein *et al.*, 2005; Guenther *et al.*, 2005; Squazzo *et al.*, 2006) (Figure 2). Unlike conventional DNA

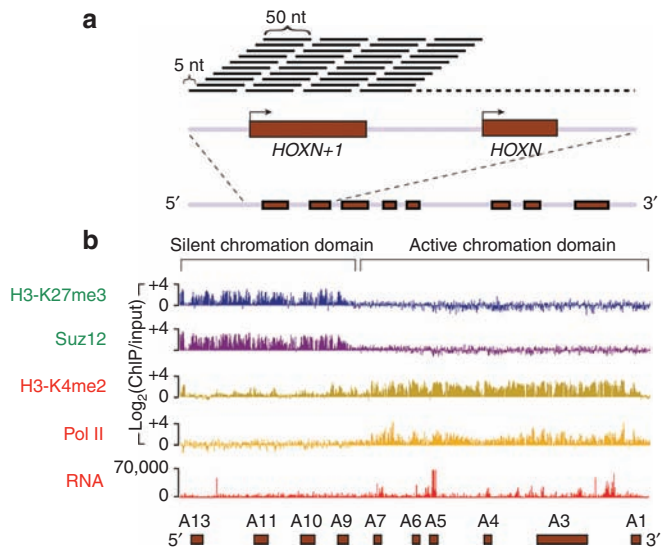

Figure 2. The epigenetic regulation of HOX expression patterns in fibroblasts. (**a**) Schematic representation of the DNA tiling design consisting of ~400,000 fifty-nucleotide probes that overlap by 45 nt such that each probe interrogates a unique 5 nt sequence. The tiling array covers all four human HOX loci and 2 Mb of control regions. Brown boxes represent specific HOX genes. (**b**) Opposing histone modifications demarcate transcriptional accessibility of HOXA locus. The top four rows show occupancy for the named protein (polycomb group protein Suz12 or RNA polymerase II (PolII)) or histone modification (H3K27me3 and H3K4me2) across ~100 kb of the HOXA locus (*x* axis), as measured by ChIP-chip. The *y* axis plots the ratio of hybridization signals of chromatin immunoprecipitation over input genomic DNA in log 2 space. The bottom row shows RNA hybridization signal over the same genomic region in linear scale (0–70,000 intensity units). Chromatin immunoprecipitation experiments were performed as described (Squazzo *et al.*, 2006).

microarrays where one probe is placed for each gene of interest, the tiling microarray employs many more probes to tile across genomic regions of interest with large overlap between adjacent probes (Figure 2a). We constructed a HOX tiling array that interrogates all four human HOX loci (HOXA, B, C, D) using 50-mer oligonucleotide probes that overlap by 45 nucleotides (nt). Thus, each 5 nt of DNA sequence is uniquely interrogated. Hybridization of RNA or chromatin immunoprecipitation on tiling microarrays (so-called ChIP-chip analysis) then reveals all of the transcribe regions or sites of chromatin modification(s) in an unbiased manner. Conventional ChIP experiments typically yield DNA fragments of 500–1,000 base pairs (bp) and hence can be interrogated by much lower resolution promoter or tiling arrays. However, interrogation of nucleosome positioning (~140 bp per nucleosome) and novel noncoding RNA (including small RNAs < 100 nt) require substantially higher resolution for their demarcation.

We and others have found evidence of broad domains of HOX chromatin modifications that demarcate active and silent regions of the HOX loci (Bernstein *et al.*, 2005; Guenther *et al.*, 2005; Squazzo *et al.*, 2006). Actively transcribed domains of HOX are marked by RNA polymerase II occupancy and histone H3 lysine 4 dimethylation in the gene-coding and intergenic regions whereas transcriptionally silent domains are marked by histone H3 lysine 27 trimethylation and Polycomb group protein occupancy, such as the Polycomb group protein subunit Suz12 (Figure 2b). These chromatin domains are regulated in a site-specific manner. Moreover, a number of noncoding RNAs are transcribed in a site-specific manner (Sessa *et al.*, 2007).

In addition to normal homeostasis, a similar genomic approach may be employed to investigate diseases of the skin. The HOX loci are also important for epidermal appendage development and regulation of cell growth. HOXC13 is required for hair outgrowth (Godwin and Capecchi, 1998). Translocation of the human Trithorax protein gene MLL, leading to activation of the HOX loci, is a frequent etiology of leukemias (Guenther *et al.*, 2005), and the Polycomb group protein gene EZH2 is frequently amplified in human epithelial tumors and in melanoma (Bracken *et al.*, 2003). For instance, studies of genome-wide occupancy sites of MLL in leukemic cells have suggested that it functions as transcriptional start site-specific histone methyltransferase (Guenther *et al.*, 2005). With the easy accessibility of skin disease tissues and cells, interrogation with tiling microarrays is a promising tool to enhance our understanding of the epigenetic regulation of regulatory genes in skin and how these epigenetic codes may be disrupted during disease.

Tissue microarray: genomics and proteomics in the context of skin architecture

Gene and tiling microarrays are useful for elucidating the genetic and epigenetic elements that differentiate cell types across the body; however, it is equally important to determine the *in vivo* anatomic localization of genes in the three-dimensional context of the skin. To address this challenge, we have constructed a ''skin diversity'' tissue microarray, where multiple skin sections are placed on a single slide to be used for *in situ* hybridization and immunohistochemistry. Our tissue microarray is comprised of 42, two-millimeter formalin-fixed, paraffin-embedded cores of skin from diverse anatomic sites and 8 internal organs such as cervix, intestine, lung, liver and bone (Figure 3a). Immunohistochemistry or RNA *in situ* hybridization can be performed on all 50 tissues in parallel, allowing unbiased and high-throughput comparison of protein or gene expression levels and localization. A potential limitation of this technology is that proteins and mRNAs present in low levels may be better visualized in frozen sections than formalin-fixed tissues, and conditions for antigen retrieval or signal amplification may need to be developed to visualize low-abundance gene products.

To illustrate the use of such a skin diversity tissue microarray, we performed RNA *in situ* hybridizations for keratin 14 and keratin 9. As expected,

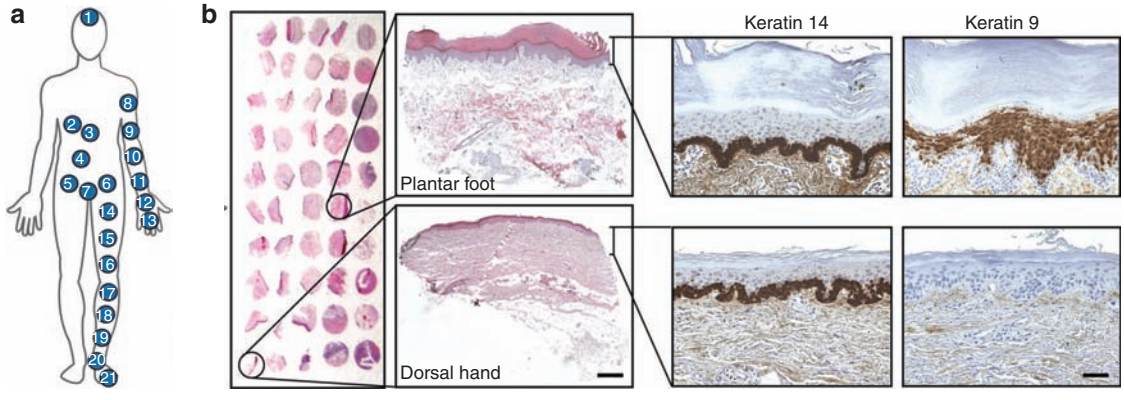

Figure 3. A human skin diversity tissue microarray. (**a**) Positional map of tissue microarray (left) comprised of 42, two-millimeter longitudinal cross-sections of skin and 8 sections of other organs, in total representing 21 unique sites of skin (blue circles). Hematoxylin and eosin staining of the tissue microarray (middle) and zoom in of cores representing plantar (top) and dorsal hand skin (bottom). Bar = 250 μm. (**b**) Skin section from plantar skin (top) and dorsal hand (bottom) following *in situ* hybridization with probes complimentary to either keratin 14 or keratin 9 mRNA. Bar = 50 μm.

we observed expression of keratin 14 in the basal layer of epidermis in skin from all anatomic sites. Keratin 9 is a suprabasal keratin of palmoplantar skin, and indeed only palmoplantar skin on our tissue microarray showed strong keratin 9 signal (Figure 3). The skin diversity tissue microarray should be useful for the discovery and validation of novel site-specific genes or signaling pathways. Again, this approach can be extended to include tissues of skin diseases to monitor the *in vivo* expression of genes perturbed in disease. The combined power of gene expression, tiling and tissue microarrays will greatly facilitate our understanding of the genes that are important in skin patterning and their roles in skin disease.

Conclusion and future challenges

The use of multiple systems biology approaches has started to paint a picture of the diversity of human dermal fibroblasts and the anatomic patterning of skin at the molecular level. This area of investigation is still at an early stage and much remains to be learned. Four challenges are likely to engross investigators in the near future. First, the transcriptional network of HOX genes in adult skin needs to be clarified. The mammalian targets of human HOX genes have eluded detection for many years owing to lack of human material and the embryonic lethality of most HOX genes in mice (Svingen and Tonissen, 2006). Primary human fibroblasts may be a tractable system to study the transcriptional network of HOX genes. Second, the epigenetic mechanisms that maintain site-specific gene expression programs remains incompletely understood. The identification of specific chromatin domains, their specific histone modifications, and associated noncoding RNAs are providing a list of candidate factors that may play a role in this regulatory program. Third, how these mechanisms of positional identity in normal skin may relate to the pathogenesis of many skin diseases with site-specific manifestations is largely unknown. Nonetheless, the history of investigative dermatology provides many examples of developmental pathways that become subverted and drive skin diseases, including diseases of epidermal adhesion, inflammation, and skin cancers. Fourth, site-specific differentiation is a dynamic process. For instance, hair cycling in many species occur in a seasonal fashion, and in mice hair cycling progresses gradually in an anterior–posterior fashion (Stenn and Paus, 2001). The use of conditional genetic approaches that can synchronize dynamic developmental programs by inducible arrest and release may aid to capture the temporal regulation of site-specific differentiation (Hutchin *et al.*, 2005; Sarin *et al.*, 2005; Estrach *et al.*, 2006). It is likely that a multifaceted approach incorporating emerging technologies will provide a wealth of information about the molecular cues involved in skin patterning and their dysregulation in skin diseases.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

We thank members of Chang lab and the Stanford Program in Epithelial Biology for discussions. This study was supported by NIAMS K08-AR050007 (H.Y.C.), Damon Runyon Cancer Research Foundation Postdoctoral Fellowship (J.L.R.), National Science Foundation Predoctoral Fellowship (J.K.W.), and Stanford Med Scholars Program and American Skin Association (H.L.). H.Y.C. is the Kenneth G. and Elaine A. Langone Scholar of the Damon Runyon Cancer Research Foundation.

REFERENCES

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120:169–81

Blanpain C, Lowry WE, Geoghegan A, Polak L, Fuchs E (2004) Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. *Cell* 118:635–48

Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K (2003) EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J* 22:5323–35

Chang HY, Chi J-T, Dudoit S, Bondre C, van de Rijn M, Botstein D *et al.* (2002) Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc Natl Acad Sci USA* 99:12877–82

Chi JT, Chang HY, Haraldsen G, Troyanskaya OG, Chang DS, Wang Z *et al.* (2003) Endothelial diversity revealed by global expression profiling. *Proc Natl Acad Sci USA* 100:10623–8

Cunha GR (1994) Role of mesenchymal–epithelial interactions in normal and abnormal development of the mammary gland and prostate. *Cancer* 74:1030–44

Dhouailly D (1973) Dermo-epidermal interactions between birds and mammals: differentiation of cutaneous appendages. *J Embryol Exp Morphol* 30:587–603

Dhouailly D (1984) Specification of feather and scale patterns. In: Malincinski G, Bryant S (eds). *Pattern formation*. New York: Macmillan Pub. Co, 581–601

Dunphy JE (1963) The fibroblast—a ubiquitous friend of the surgeon. *N Engl J Med* 268: 1367–77

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–8

Estrach S, Ambler CA, Lo Celso C, Hozumi K, Watt FM (2006) Jagged 1 is a beta-catenin target gene required for ectopic hair follicle formation in adult epidermis. *Development* 133:4427–38

Gabbiani G, Rungger-Brindle E (1981) The fibroblast.In: Glynn LE (ed). *Handbook of inflammation tissue repair and regeneration*. Amsterdam: Elsevier/North Holland Biomedical Press, 1–50

Godwin AR, Capecchi MR (1998) Hoxc13 mutant mice lack external hair. *Genes Dev* 12:11–20

Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, Canaani E *et al.* (2005) Global and Hox-specific roles for the MLL1 methyltransferase. *Proc Natl Acad Sci USA* 102: 8603–8

Hutchin ME, Kariapper MS, Grachtchouk M, Wang A, Wei L, Cummings D *et al.* (2005) Sustained Hedgehog signaling is required for basal cell carcinoma proliferation and survival: conditional skin tumorigenesis recapitulates the hair growth cycle. *Genes Dev* 19: 214–23

Jahoda CA, Horne KA, Oliver RF (1984) Induction of hair growth by implantation of cultured dermal papilla cells. *Nature* 311: 560–2

Jahoda CA (1992) Induction of follicle formation and hair growth by vibrissa dermal papillae implanted into rat ear wounds: vibrissa-type fibres are specified. *Development* 115: 1103–9

Levinson R, Mendelsohn C (2003) Stromal progenitors are important for patterning epithelial and mesenchymal cell types in the embryonic kidney Semin. *Cell Dev Biol* 14:225–31

Millar SE (2002) Molecular mechanisms regulating hair follicle development. *J Invest Dermatol* 118:216–25

Normand J, Karasek MA (1995) A method for the isolation and serial propagation of keratinocytes, endothelial cells, and fibroblasts from a single punch biopsy of human skin. In vitro *Cell Dev Biol Anim* 31:447–55

Rendl M, Lewis L, Fuchs E (2005) Molecular dissection of mesenchymal–epithelial interactions in the hair follicle. *PLoS Biol* 3:e331

Ringrose L, Paro R (2004) Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* 38:413–43

Rinn JL, Bondre C, Gladstone HB, Brown PO, Chang HY (2006) Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet* 2:e119

Sarin KY, Cheung P, Gilison D, Lee E, Tennen RI, Wang E *et al.* (2005) Conditional telomerase induction causes proliferation of hair follicle stem cells. *Nature* 436:1048–52

Schedl A, Hastie ND (2000) Cross-talk in kidney development. *Curr Opin Genet Dev* 10: 543–9

Schmitt S, Paro R (2006) RNA at the steering wheel. *Genome Biol* 7:218

Schwann T (1847) *Microscopical researches into the accordance in the structure and growth of animals and plants*. London: The Sydenham Society, 1–268

Sessa L, Breiling A, Lavorgna G, Silvestri L, Casari G, Orlando V (2007) Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. *RNA* 13: 223–39

Shannon JM, Hyatt BA (2004) Epithelial–mesenchymal interactions in the developing lung. *Annu Rev Physiol* 66:625–45

Sneddon JB, Zhen HH, Montgomery K, van de Rijn M, Tward AD, West R, *et al.* (2006) Bone morphogenetic protein antagonist gremlin 1 is widely expressed by cancer-associated stromal cells and can promote tumor cell proliferation. *Proc Natl Acad Sci USA* 103:14842–7

Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, Jang SW *et al.* (2006) Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res* 16:890–900

Stenn KS, Paus R (2001) Controls of hair follicle cycling. *Physiol Rev* 81:449–94

Svingen T, Tonissen KF (2006) Hox transcription factors and their elusive mammalian gene targets. *Heredity* 97:88–96

Wolpert L (1969) Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* 25:1–47

Yamaguchi Y, Itami S, Tarutani M, Hosokawa K, Miura H, Yoshikawa K (1999) Regulation of keratin 9 in nonpalmoplantar keratinocytes by palmoplantar fibroblasts through epithelial–mesenchymal interactions. *J Invest Dermatol* 112:483–8